CESifo Working Papers
www.cesifo.org/wp

Interpreting Time Horizon Effects in Inter-Temporal Choice

Thomas Dohmen
Armin Falk
David Huffman
Uwe Sunde

CESIFO WORKING PAPER NO. 3750
CATEGORY 13: BEHAVIOURAL ECONOMICS
FEBRUARY 2012

CESifo *Working Paper No. 3750*

Interpreting Time Horizon Effects in Inter-Temporal Choice

Abstract

We compare different designs that have been used to test for an impact of time horizon on discounting, using real incentives and two representative data sets. With the most commonly used type of design we replicate the typical finding of declining (hyperbolic) discounting, but with other designs find constant or increasing discounting. As a whole, the data are not consistent with any of these usual candidate discounting assumptions, and they also imply a violation of transitivity. The results have implications for interpreting previous evidence, and pose an important puzzle for understanding inter-temporal choice.

JEL-Code: D010, D900, D030, E210.

Keywords: time preference, hyperbolic discounting, self-control, dynamic inconsistency, intransitivity.

Thomas Dohmen
Research Centre for Education and the Labour Market (ROA)
Maastricht University
Maastricht / The Netherlands
t.dohmen@maastrichtuniversity.nl

Armin Falk
Institute for Empirical Research in Economics
University of Bonn
Bonn / Germany
armin.falk@uni-bonn.de

David Huffman
Department of Economics
Swarthmore College
Swarthmore / PA / USA
dhuffma1@swarthmore.edu

Uwe Sunde
Swiss Institute for Empirical Economic Research
University of St. Gallen
St. Gallen / Switzerland
uwe.sunde@unisg.ch

February 16, 2012

1 Introduction

Understanding inter-temporal decision making is crucial for economics, and choice experiments have been an important tool in this undertaking. Early experiments in psychology had a particularly important impact on economics. These showed that exogenously varying the time horizon affected inter-temporal choice, in a way consistent with declining rather than constant discounting (for a survey see Frederick et al., 2002). This evidence provided part of the motivation for new economic models incorporating hyperbolic, or quasi-hyperbolic, discount functions (e.g., Loewenstein and Prelec, 1992; Laibson, 1997; O'Donoghue and Rabin, 1999), with important implications in terms of the possibility of dynamic inconsistency and self-control problems. Similar choice experiments have also been used by economists as a way to potentially measure the extent of self-control problems among different groups of people, or predict economic outcomes thought to be related to self-control (e.g., Sutter et al., 2010; Meier and Sprenger, 2010).

Taking into account more recent evidence, however, it is less clear that time horizon effects in experiments are in fact capturing the shape of an underlying discount function. While initial studies did find declining discounting, this was predominantly (though not exclusively) using one type of design.[1] Other studies using different but theoretically equivalent designs have found constant discounting, increasing discounting, or some other type of choice pattern, on average (e.g., Anderhub et al., 2001; Read, 2001; Read and Roelofsma, 2003; Rubinstein, 2003; Read et al., 2005; Harrison et al., 2005; Zauberman et al., 2009; Ebert and Prelec, 2007; Benhabib et al., 2010; Andrioni and Sprenger, 2010; Andersen et al., 2010; Takeuchi, 2010; Andersen et al., 2011; Sutter et al., 2010; Halevy, 2011). Also, studies relating experimental measures of dynamic inconsistency to life outcomes have found mixed results, and other mechanisms besides self-control problems could possibly explain the relationships that are observed.[2] Thus, it is not clear what discount-

[1] Frederick et al. (2002) survey ten studies that compare discounting over overlapping time horizons of different lengths (what we call an "overlapping design" below), and find behavior consistent with declining discounting. They also find the same pattern comparing *across* roughly thirty studies that use either short or long time horizons. By contrast, only four studies use a different design, involving comparing non-overlapping time horizons of equal lengths (what we call a "shifted design" below), and find evidence that can be interpreted as declining discounting. Some of these latter studies also have some unusual design features, which make comparison to subsequent experiments conducted by economists more difficult.

[2] In some cases studies have found no difference in outcomes based on usual measures of dynamic inconsistency (Sutter et al., 2010), or mixed evidence depending on the outcome being studied (Meier and

ing assumption is a good description of the evidence as a whole, or if any discounting model explains the data. Resolving the controversy has been difficult, however, partly due to differences across studies in terms of the approach used to measure time horizon effects.

This paper provides new evidence on how time horizon affects inter-temporal choice, in a design that compares multiple measurement approaches within the same framework. This allows investigating whether seeming support for a given discounting assumption varies systematically with the type of measure used. Also, we can look at the ability of discounting models to explain the data from a more holistic point of view. Our design is also powerful because it combines several other features. We use two relatively large data sets, with 500, and 1,500 subjects respectively, each of which was drawn so as to be representative of the adult population in Germany. The use of representative samples helps establish whether time horizon effects are specific to certain segments of the population, such as students, or are more pervasive and thus more important. We use real incentives, in contrast to some studies that use hypothetical incentives, and credibility of our incentives is particularly high because subjects are in a long-term relationship with the agency conducting the experiments. We vary stake sizes, and also method of payment delivery, across different groups of subjects, which allows us to assess robustness to such design variations. Finally, we have detailed information on life outcomes that are often thought to reflect self-control problems, allowing us to contribute evidence on how these relate to experimental measures of time horizon effects.

We designed the experiment to explicitly compare three different approaches to testing discounting assumptions. All of these build on the workhorse, "price-list" framework that is the industry standard for economists: Subjects make a series of choices between smaller, sooner and larger, later monetary rewards. The most commonly used approach, which we denote the "overlapping design" (OD), involves measuring discounting over two or more overlapping time horizons for the same individual, where all time horizons start on the same date. For example, one of our OD measures involves comparing discounting between the immediate future, denoted 0, and 6 months, to discounting between 0

Sprenger, 2008; Meier and Sprenger, 2010; Burks et al., 2011) or gender of the subject (Ashraf et al., 2006; Bauer et al., 2010). With a more parametric approach, Harrison et al. (2010) find that smokers and non-smokers are "no more or no less dynamically inconsistent than each other" (pp. 716). Chabris et al. (2008) find a relationship between discounting measures and outcomes, but using the level of the discount rate rather than measures of sensitivity to time horizon.

and 12 months.[3] Studies using this design often find a pattern consistent with declining (hyperbolic) discounting (Frederick et al., 2002). Another approach, which we denote the "shifted design" (SD), involves two non-overlapping time horizons of the same length, where the start date of one time horizon is shifted forward in time. For example, one of our SD measures compares discounting between 0 and 6 months to discounting between 6 months and 12 months. While the shifted design corresponds to the commonly cited "apple" thought experiment by Thaler (1981), it has actually been less commonly used than the overlapping design.[4] Studies that implement this approach sometimes find that behavior is consistent with declining discounting (e.g., Kirby and Herrnstein, 1995, and others), but in other cases find constant or even increasing discounting (e.g, Anderhub et al., 2001; Sutter et al., 2010). We also designed the experiment to allow what we call the "overlapping, shifted design" (OSD). This involves comparing discounting over a long time horizon to discounting over a shorter time horizon, which overlaps but is shifted forward so as to start later in time. For example, we compare discounting between 0 and 12 months to discounting between 6 months and 12 months. This latter approach was initially tried by Baron (2000) and Read (2001), and yielded little support for declining discounting.

Importantly, all of these approaches to distinguishing between discounting models rely on a set of maintained assumptions. If these assumptions do not hold, models typically predict constant discounting behavior in the context of the experiments, regardless of the shape of the discount function.[5] In this case, however, there is still a testable prediction for the broader class of models, namely that behavior should be in line with constant discounting. We assess the ability of different models to explain the data, both with and without the usual maintained assumptions.

The main stylized fact from our analysis is that time horizon matters for intertemporal choice, in a way that is hard to explain with standard discounting models: People

[3] Use of 0 to denote the immediate payments is a slight abuse of notation, because payments were never truly immediate; they arrived in the immediate future, typically two days from the date of the experiment. We discuss the reasons for this (commonly used) front end delay feature of the design in more detail below.

[4] The thought experiment involves a choice between 1 apple today or 2 apples tomorrow, and between 1 apple in a year or 2 apples in a year and a day. Both choices are made today. Declining discounting can explain why people might potentially choose 1 apple in the early horizon, but plan to choose 2 apples in the later horizon.

[5] For example, it has typically been assumed that people treat monetary payments as consumption. We discuss why relaxing this assumption means that constant and non-constant discounting cannot be distinguished, because both predict constant discounting behavior in the experiment context.

are more impatient for short time horizons than longer time horizons, inconsistent with constant discounting, but are relatively insensitive to when a given time horizon starts, contrary to non-constant discounting. For example, people are more impatient from 0 to 6 months than 0 to 12 months, but similarly impatient for 6 to 12 months compared to 0 to 6 months. This pattern is also inconsistent with a broader class of models, in that it implies a violation of transitivity.[6] The same qualitative pattern is observed in our second data set, despite using different parameters, time horizon lengths, and starting dates. The pattern is also robust in that it holds for different sub-populations that we consider (by gender, age, education, cognitive ability, and financial situation). To make sure that the aggregate patterns are representative of individual level behavior, rather than reflecting a misleading composition effect from heterogeneous types, we conduct individual-level analysis. The modal choice pattern for individuals is exactly as observed at the aggregate level, and only a few individuals can be classified as consistent with standard discounting assumptions. In conservative robustness checks that allow for substantial "trembling hand errors", the relative majority still violate transitivity.

A first implication of the results is that there is an important puzzle for understanding inter-temporal choice. The types of experiments we conduct have traditionally been a yardstick for assessing different modeling assumptions, as they involve incentivized choices, and relatively simple inter-temporal trade-offs. The observed sensitivity of discounting to time horizon length has potentially profound implications for a wide range of economic choices. Yet, discounting models, regardless of whether discounting is constant or non-constant, have trouble fully explaining these effects. This points to an opportunity to improve models of inter-temporal decision making, in a way that may prove useful both in the context of experiments and also in the field. We discuss some potential directions for future theoretical work at the end of the paper.

A second main implication concerns the interpretation of previous empirical research on time horizon effects. Our findings offer a potentially useful lens through which to view "mixed evidence" across previous studies: The particular method variance we observe maps, broadly speaking, onto the patterns of inconsistent results. We reliably find declining discounting with OD measures, similar to most previous studies using this approach,

[6] See Roelofsma and Read (2000) for some of the first evidence that inter-temporal choices can violate transitivity.

and less evidence for declining discounting using other types of measures, similar to many (but not all) previous studies using SD or OSD approaches. Thus, the original focus on OD designs in the literature may have contributed to the perception that declining discounting could explain time horizon effects. Relatedly, our results raise questions about how to interpret evidence from the many studies that use only one or perhaps two types of time horizon comparisons. These may seem to support one particular discounting assumption or another, when in fact a more comprehensive design might reveal pervasive intransitivity and inconsistency with any standard discounting model. Another implication concerns the use of experimental measures to identify self-control problems. One potential explanation for the mixed success of previous studies in predicting life outcomes related to self-control could be attenuation bias: without a design involving multiple time horizon comparisons, there may be a "biased" classification of discounting types, such that many individuals who are classified as, say, constant discounters may in fact be intransitive. However, given our finding that the majority of time horizon effects cannot be explained by non-constant discounting, an alternative explanation for the weak results is that some other mechanism besides self-control problems drives time horizon effects. In our concluding discussion, we report results on how our experimental measures relate to life outcomes, using a specification that addresses attenuation bias, and conclude in favor of the latter explanation.

Our results are complementary to another strand of the literature in psychology, which has questioned whether time horizon effects reflect discounting. Previous studies (e.g., Barron, 2001; Read, 2001; Read and Roelofsma, 2003; Read et al., 2005; Zauberman et al., 2009) found, as in our study, "subadditivity" of discounting on average, such that discounting is more extreme when measured using sub-intervals.[7] For the most part, however, these studies have had hypothetical incentives, and have used college student subjects.[8] Our study is different in that we use representative samples of adults, combined with incentivized experiments and highly credible payments. We consider the ability of

[7] The tendency for people to be more impatient for both 0 to 6 months, and 6 to 12 months, than 0 to 12 months, implies subadditivity: The total discounting from compounding the discount factors for the two sub-intervals is less than the discount factor obtained from the longer time horizon measure. Maintaining usual assumptions of separability and time stationarity, additivity is a prediction of all standard discounting models, regardless of the shape of the discount function.

[8] Exceptions are Experiment 3 in Read (2001), where one of 20 student subjects was randomly selected to be paid, and Roelofsma and Read (2005), where one of 2,000 participants in a (non-representative) internet experiment was randomly selected to receive real money.

various models to explain the behavior we observe, under different possible assumptions. We are able to show a pervasive violation of discounting models and transitivity in the general population, and show the robustness of this result across different sub-populations. Finally, we relate the experimental measures to real economic outcomes.

The findings of this paper also add to a literature in economics, on time discounting. For example, Harrison et al. (2002) were the first to study discounting experiments in a representative sample, and Andersen et al. (2010; 2011) also do so.[9] Different from our study, both focus on OD type designs. Bisin et al. (2010) focus on OD comparisons in the lab, and Andrioni and Sprenger (2010) use multiple types of time horizons in the lab, but do not explicitly consider method variance or intransitivity. Our paper adds an important additional dimension, showing how a design with multiple types of time horizon comparison can deliver a quite different perspective on the nature of inter-temporal choice and ability of various models to explain time horizon effects. Some recent studies have also deviated from the typical approach of eliciting all choices, about the present and about the future, at one point in time; returning to subjects multiple times, they test whether the pattern of declining discounting implied by choices made in the initial interview actually translate into the predicted direction of choice reversal (Harrison et al., 2005; Airoldi et al., 2011; Halevy, 2011), and so far conclude in the negative. Our approach is complementary in that it shows that even the initial choices are not consistent with discounting predictions, and thus helps explain why returning to subjects in the future does not yield behavior consistent with the predictions of discounting models.

The rest of the paper is organized as follows. Section 2 describes the data sets, treatments, and behavioral predictions. Section 3 presents the main results, and discusses ability of different models to explain the stylized facts. Section 4 concludes with a discussion of directions for future research, and potential implications for policy interventions intended to target self-control problems.

[9] Their approach is complementary and has different strengths. In particular, they use a more parametric methodology, and additional experimental measures to calibrate concavity of utility.

2 Design of the Experiment

2.1 Data collection and experimental procedures

Our analysis uses two data sets. One data set, which we call the SOEP data, involves a sub-sample of participants in the German Socio-Economic Panel (SOEP), a large panel data set for Germany (for a detailed description of the SOEP see Schupp and Wagner, 2002; Wagner et al., 2007). The second data set is the SOEP Cross Sectional Study 2005, which we call CSS data for short. This data set involves a separate sample of individuals, which was collected by the SOEP administration as part of the annual process of "pretesting" questions for potential use in the SOEP survey; the data set is just a cross-section rather than a panel like the actual SOEP. The CSS data, and the SOEP data, were collected during Spring of 2005, and Spring and Summer of 2006, respectively. Collection in each case was done by the same professional surveying company that administers the SOEP every year. Sampling for each data set was done according to the same procedure used to generate the SOEP sample, and individuals were visited by interviewers in their own homes.[10] Both the CSS and SOEP samples were constructed so as to be representative of the adult population, age 17 and older, living in Germany.[11] In total the CSS data include 500 subjects who participated in the inter-temporal choice experiments, and the SOEP data include 1,503 such subjects.

Participants in our studies went through a computer assisted personal interview (CAPI) conducted with a laptop. The interview consisted of two parts. First, subjects answered a detailed questionnaire. The items in the questionnaire were presented in the standard format used by the SOEP. Topics included demographic characteristics, financial situation, health, and attitudes. The questionnaire also included two brief tests of cognitive ability. The full questionnaire, in German and translated into English, is available upon request. At the end of the questionnaire, subjects were invited to participate in the second part of the interview, which consisted of a paid experiment.

The first step in the experimental procedure involved the experimenter presenting

[10] For each of 179 randomly chosen primary sampling units (voting districts), an interviewer was given a randomly chosen starting address. Starting at that specific local address, the interviewer contacted every third household and had to motivate one adult person aged 17 or older to participate. For a detailed discussion of the random walk method of sampling see Thompson (2006).

[11] Respondents had to turn 18 during the year of the interview to be eligible.

subjects with some example choices. The experimenter explained the types of choices that the subject would make, and how payment would work. In particular, subjects were informed that the experiment would involve multiple choices, that one choice situation would be randomly selected after all choices had been made, and that the choice they made in this situation would potentially be relevant for their payoff. Subjects knew that at the end of the experiment a random device would determine whether they were actually paid, with the probability of being paid equal to 1/7 in the CSS, and 1/9 in the SOEP data. This procedure gives subjects an incentive to choose according to their true preferences in each choice situation, and thus is incentive compatible. After explaining the nature of the experiment and the rules for payment, the experimenter asked subjects whether they were willing to participate. Subjects who agreed to participate were given further instructions, and then allowed to ask questions. Once there were no more questions, the experiment began, and subjects were asked to make their actual choices. An example of the script and instructions used in the experiments is presented in Appendix A.2 below, translated from German into English.

Our experiments were designed to give a measure of the annual internal rate of return (IRR) needed to induce an individual to wait, for a given time horizon. A time horizon Tst is defined by starting date s, and ending date t. For a given horizon an individual chose between an early payment, X_s, available at the start of the horizon, or a larger, later payment, Z_t, available at the end of the horizon. For example, in horizon T012, the early payment is at time 0, and the later payment is 12 months in the future. In all choices for a given horizon, the amount of X_s was held constant, but the later payment, Z_t, was larger in each subsequent choice. For most time horizons, the value of Z_t in the first choice was calibrated to be consistent with an annual IRR of 2.5 percent, assuming semi-annual compounding, and each subsequent value of Z_t implied an additional 2.5 percentage point increase in the annual rate of return, up to a maximum of 50 percent.[12] In one treatment we implemented choices with a coarser measurement of the IRR, in steps of 5 percentage points, but allowed measuring annual IRRs as high as 100 percent. We obtain an incentive compatible measure of impatience for a given time horizon by observing the value of Z_t, or

[12] We chose semi-annual compounding of the annual interest rate because this is a natural compromise between the two types of compounding German subjects are most familiar with: quarterly compounding on typical bank accounts, and annual reports on the rate of return from savings accounts, pension funds, or stock holdings. Using semi-annual compounding also helps avoid prominent round numbers in the choices, which could potentially influence switching choices.

equivalently the annual IRR, needed to induce the individual to wait. Actually, we obtain upper and lower bounds for the annual IRR, separated by 2.5 percentage points, due to the discrete variation in late payment amounts; in our analysis we focus on lower bounds. Across treatments, we varied s and t, and the amounts X_s and Z_t.

During the experiment subjects were presented with the different choices, one at a time, on the computer screen. The first time that a subject switched from the early to the delayed payment, the subject was asked whether he or she also preferred to wait for any larger payment. All subjects agreed, and as a consequence, there are no non-monotonic choices in the data. It has been shown previously that multiple-switching between early and late payment amounts is associated with lower cognitive ability (Burks et al., 2009); our procedure was intended to help mitigate potential confusion among subjects about the structure of the choice problem, and reduce frequency of non-monotonic choices, by re-iterating the tendency for the late payment to only increase across further choices.[13]

Subjects knew that one row would be randomly selected at the end of the experiment, and that their decision in that row could be relevant for their payoff. Subjects also knew that all payments would be sent by mail following the interview. For the CSS data set, all payments were mailed on the day after the interview and thus would arrive within at most two days due to the well-known two-day guarantee for delivery by the German postal service.[14] Certificates for immediate payments could be cashed immediately, once they arrived, while certificates for payments in the future could only be cashed at the specified time. For the SOEP data, payments were also sent by mail, but the timing of the mailing reflected the timing of the payments, i.e., immediate payments were mailed immediately after the experiment and arrived within two days, while later payments were mailed punctually on the corresponding later date. This variation in procedures provides a potentially useful check on whether timing of payment deliveries matters for results, in

[13] One concern with this type of procedure might be that it tends to "lock-in" mistakes, if subjects for some reason tend to make a mistake the first time they choose the later payment, and then do not want to tell the experimenter that they have changed their mind. A similar concern would apply to a standard "titration" procedure that is often used in inter-temporal choice experiments, where the choices offered to subjects depend on past choices. Even if mistakes were locked-in, however, to the extent that this affects all measures it should not matter for our conclusions, because they are based on differences across measures. Furthermore, we find very similar results regardless of cognitive ability, suggesting that our results are not driven by those who might be more prone to mistakes. Notably, studies that use some other approach, and end up with more non-monotonic choices, are put in the position of either dropping (non-randomly) a part of the data, or incorporating non-monotonic choices in an ad hoc way.

[14] The Deutsche Post is highly successful at achieving an explicit goal to delivery mail within two days, for packages with origin and destination within Germany.

particular due to credibility issues.

In the literature on inter-temporal choice experiments, it is desirable to have equal credibility of early and late payments, and our design has several features that arguably achieve this goal as well or better than previous studies. Equal credibility is desirable in order to prevent, e.g., that early payments enjoy relatively higher credibility, and thus choices are distorted in the direction of overstating impatience, or even overstating the degree of declining discounting.[15] A first feature of our design that addresses credibility is the "front-end delay" (see Coller and Williams, 1999), in the sense that all payments arrived by mail after the experiment. Thus, early payments did not have any special credibility arising from being paid during the experiment session itself. This is a standard procedure in discounting experiments to help make early and late payments equally credible. Second, experiments for both data sets were conducted by the professional agency used by the SOEP, which is highly credible and well-known because of its role in conducting election polls for German public television. Interviewers also left their contact details at the end of the experiment, making it easy for subjects to contact the institute. There were no reports, from any of the interviewers, about subjects expressing concerns regarding credibility of payments. Third, all participants in the SOEP data were members of the SOEP panel itself; unlike participants in most inter-temporal choice experiments, these individuals were in a long-term relationship with the individual surveyor conducting the experiment. Thus, perceived credibility of payments should have been very strong regardless of timing. Fourth, we find the same qualitative results across our two data sets, in terms of how behavior varies with time horizon length and timing, despite using different procedures for the timing of payment deliveries. This suggests that credibility concerns linked to timing of payment delivery do not drive the results.[16]

[15] As pointed out by Andrioni and Sprenger (2010) and others, unequal credibility of payments could generate the appearance of declining discounting in SD comparisons, because the earlier time horizon includes an immediate, especially credible payment while the later horizon does not. This is particularly likely if the early payment is actually received during the experimental session. For a series of reasons discussed in the text, however, credibility is high in our design, and thus it is unlikely that unequal credibility could explain time horizon effects in our particular data set. Furthermore, it would not explain patterns that resemble declining discounting in OD comparisons, constant discounting in SD comparisons, or increasing discounting in OSD comparisons, all of which we find. Thus, while unequal credibility of payments could in some settings generate spurious declining discounting, this does not seem to explain the time horizon effects we observe.

[16] For the SOEP data, payments were mailed at the time that they became cashable, and thus credibility depended on the perception that the agency would in fact deliver the payments as promised. For all of the reasons discussed above, there should have been no credibility concerns about payments from the agency in general, as well as no special concerns about relative credibility of late payments. The

A final note on the design concerns the length of the front-end-delay. Many previous studies that test discounting assumptions have had front-end-delays ranging from one day to as long as one month (e.g., Meier and Sprenger, 2010; Harrison et al., 2002). While useful for equalizing credibility, a potential drawback of such a delay is reduced "immediacy" of early payments. Indeed, some non-constant discounting models, which assume a discrete drop in discount rates between the present and future, are based on the intuition that present-bias reflects psychological processes that respond to immediate rewards. To the extent that the front-end-delay eliminates immediacy of early payments, such a design feature may "miss" present-bias. Unfortunately, the verdict is still out on how immediate payments would need to be, in order to trigger the aforementioned psychological processes. Qualitatively, results actually appear quite similar across designs that do and do not have same-day payments, e.g., patterns consistent with declining discounting have been found in both.[17] Making a trade-off between credibility and immediacy, we chose the shortest possible front-end delay compatible with avoiding a same-day credibility problem.[18] Importantly, models where present bias occurs within the window of the front-end-delay still make clear testable predictions in the experiment, namely constant IRRs across time horizons.

2.2 Treatments

Table 1 summarizes the various treatments. As shown in the table, the CSS data involved 500 subjects participating in the experiments, with three different measures of annual

procedure used for the CSS data did, however, potentially allow for a different type of credibility concern: the agency mailed all payments immediately regardless of when they became cashable, so individuals might worry that they themselves would be more likely to lose later payments before they become cashable (although the money on the line should have provided an incentive not to lose payments). We find the same qualitative results using the SOEP data as in the CSS data, which suggests that a potential fear of losing late payments does not drive the results in the latter case.

[17] It is quite intuitive that the present is today, and the future is tomorrow. On the other hand, one could think that immediacy is actually about having a consumption opportunity literally in one's hand, and that even a delay of ten minutes might already be too long for rewards to feel immediate. Almost without exception studies have had a delay of at least some hours before early rewards are usable, so rewards are not truly immediate (e.g., Burks et al., 2009). One exception is McClure et al. (2007), where rewards are squirts of juice over the course of minutes.

[18] Mainly for the purposes of concreteness, so as to avoid having to say 1 or 2 days repeatedly, the experimental instructions told subjects that immediate rewards would be referred to as being received "Today"[quotes included]. At the same time, the instructions were very clear that all rewards would arrive after the experiment by post, and that: "Today means you can cash the check you receive by post immediately". Ultimately, we see little evidence that this wording lead to differential behavior in early versus later time horizons. For example, observed impatience is similar regardless of whether the time horizon involves early payments "Today", or involves early payments in 12 months.

IRR for each subject. The measures come from three different time horizons, 0 to 6 months (T06), 0 to 12 months (T012), and 6 months to 12 months (T612). The order of the treatments in the CSS data was randomized across individuals. The early payment was always 100 Euros, and the largest delayed payment always implied an annual IRR of 50 percent (compounded semi-annually) for waiting the specified length of time. If individuals never chose the later payment, their IRR was right-censored, and coded as having a (lower-bound) value of 52.5 percent.

The SOEP data differ in that there are only two treatments for each individual (see Table 1), and treatments also varied across sub-samples. For a first sub-sample of 490 individuals, impatience was measured for 0 to 6 months (T06) and 0 to 12 months (T012), giving two measures of annual IRR for each subject. The second sub-sample of 487 were asked about 0 to 1 month (T01) and 0 to 12 months (T012). For the third sub-sample of 526 we measured discounting for 0 to 1 month (T01b) and 12 to 13 months (T1213). The measures for the third sub-sample were different, because IRRs were measured in steps of 5 percent rather than 2.5 percent, and the upper-bound IRR in each horizon was 105 percent rather than 52.5 percent. We denote the one-month measure in this sub-sample T01b, to distinguish it from T01 in the second sub-sample. Order was predetermined in the SOEP data: for the first two sub-samples, the T012 measure was always elicited first; for the third sub-sample, T01b was elicited first. A random device on the computer selected whether an individual was assigned to the first, second, or third sub-sample experiments. In the SOEP data the early payment was always 200 Euros, and thus stakes were higher than in the CSS, even accounting for the lower payment probability of 1/9 rather than 1/7.[19]

2.3 Behavioral Predictions

The type of experimental design we implemented has been used extensively to distinguish between constant and non-constant discounting models, based on a certain set of maintained assumptions. We first derive predictions under the usual assumptions, and then discuss how predictions change if the assumptions are relaxed. Generally, relaxing these

[19] The relatively large nominal values involved in the experiment help mitigate distortions due to subjects rounding delayed payment amounts up to the nearest dollar. See Andersen et al. (2011) for a discussion of this issue.

assumptions makes models predict constant discounting in the context of the experiment, regardless of the shape of the discount function. In this case, however, the broader class of models still makes a testable prediction, namely invariance of discounting to time horizon.

We illustrate the predictions by considering, without loss of generality, an example with three time horizons, T06, T012, and T612, corresponding to the structure of the CSS data. The early payment is always 100 for each horizon, compounding occurs once every 6 months, and for simplicity a period is assumed to be 6-months long. Assume for now that early payments at time 0 are literally available on the day of the experiment, when preferences are measured. Also, adopt the usual maintained assumptions that utility is locally linear, that people treat monetary payments as consumption, that utility is additively separable over time and time stationary, and that the discount function is multiplicatively separable.

When making decisions in T06, T012, and T612, subjects decide for the early or late payment depending on whether or not the offered *annual* rate of return r in a given choice is sufficiently attractive to induce waiting. Thus, decisions involve the following comparisons:

$$\left(1+\frac{r}{2}\right)100 \lesseqgtr Z(6); \qquad \left(1+\frac{r}{2}\right)^2 100 \lesseqgtr Z(12); \qquad \left(1+\frac{r}{2}\right)100 \lesseqgtr Z(12)$$

Where $Z(6)$, $Z(12)$, and $Z(12)$ denote later payments for each time horizon, available at 6 months, 12 months, and 12 months from the present, respectively. We denote by $Z(6)^{T06}$, $Z(12)^{T012}$, and $Z(12)^{T612}$ the lowest late payment such that the individual prefers to wait for the corresponding time horizon. These observed amounts establish the points of indifference for each horizon, and define the internal rates of return:

$$\left(1+\frac{IRR^{T06}}{2}\right)100 = Z^{T06}; \quad \left(1+\frac{IRR^{T012}}{2}\right)^2 100 = Z^{T012}; \quad \left(1+\frac{IRR^{T612}}{2}\right)100 = Z^{T612} \tag{1}$$

Note that from now on we omit the time argument for delayed payments, for ease of exposition. We also neglect the fact that the delayed payment is actually a discrete variable in the experiment, and thus that we can infer only a range for the IRR; this has no consequences for the qualitative predictions, and eases exposition.[20] Solving for IRRs

[20] The lowest delayed payment that is preferred establishes an upper bound for the IRR, while the largest delayed payment that is not preferred establishes the lower bound. One can think of the predictions as being derived based on lower bounds everywhere (or equivalently in terms of upper bounds).

yields:

$$IRR^{T06} = 2\left(\frac{Z^{T06}}{100} - 1\right); \quad IRR^{T012} = 2\left(\left(\frac{Z^{T012}}{100}\right)^{\frac{1}{2}} - 1\right); \quad IRR^{T612} = 2\left(\frac{Z^{T612}}{100} - 1\right) \tag{2}$$

We now consider how different assumptions about time preference affect predictions for IRRs.

2.3.1 Constant discounting

In the case of constant discounting, an individual is indifferent between the early and delayed payments in T06, T012, and T612 when

$$(1+\frac{\rho}{2})100 = Z^{T06}; \qquad (1+\frac{\rho}{2})^2 100 = Z^{T012}; \qquad (1+\frac{\rho}{2})100 = Z^{T612}, \tag{3}$$

where ρ is the constant rate of time preference. As this is the same as condition (1), it follows directly that a constant discounter chooses Z^{T06}, Z^{T012}, and Z^{T612} such that the measured IRR is invariant with respect to time horizon: $IRR^{T06} = IRR^{T012} = IRR^{T612} = \rho$.

2.3.2 Declining and increasing discounting

In the case of declining discounting, there are different discount rates, ρ_1 and ρ_2, for periods 1 and 2 (first 6 months and second 6 months), such that $\rho_1 > \rho_2$. In this case the points of indifference in T06, T012, and T612 are given by

$$(1+\rho_1)100 = Z^{T06}; \qquad (1+\rho_1)(1+\rho_2)100 = Z^{T012}; \qquad (1+\rho_2)100 = Z^{T612}. \tag{4}$$

Substituting into (2) shows that, if choices are generated by this model, measured IRRs will have the form

$$IRR^{T06} = 2((1+\rho_1)-1); \quad IRR^{T012} = 2(((1+\rho_1)(1+\rho_2))^{\frac{1}{2}}-1); \quad IRR^{T612} = 2((1+\rho_2)-1). \tag{5}$$

Given $\rho_1 > \rho_2$ this implies $IRR^{T06} > IRR^{T012} > IRR^{T612}$. Intuitively, impatience should be greatest in T06 with declining discounting, because it includes the present and extends the least far into the future. Behavior in T612 should be the most patient, because it excludes the present and only concerns payments relatively far into the future. Behavior in T012 should be in-between, as it includes the present but also extends substantially into the future. An analogous argument establishes that with increasing discounting, the model predicts the opposite ranking of IRRs by time horizon, $IRR^{T06} < IRR^{T012} < IRR^{T612}$.

2.3.3 Predictions with different/weaker assumptions

Relaxing the assumption that early payments arrive immediately may matter for the predictions of some types of non-constant discounting models. Specifically, the fact that early payments actually arrive up to two days in the future matters for models where there are discrete changes in the discount rate during those two days, followed by constant discounting thereafter. A particularly important example is the quasi-hyperbolic model (e.g., Phelps and Pollak, 1968; Laibson, 1997; O'Donoghue and Rabin, 1999), in the case that the period is assumed to be less than two days (the model is silent about the correct period assumption). Such a model could imply a high discount rate between today and tomorrow, and then a constant discount rate between any two future periods.[21] Given the experimental design, this type of model thus predicts $IRR^{T06} = IRR^{T012} = IRR^{T612}$, the same as with constant discounting, i.e., invariance of IRRs with respect to time horizon is needed for the model to explain the data. If the present bias is assumed to extend more than 2 days into the future, but less than 6 months, then the quasi-hyperbolic model makes the same predictions as models with hyperbolic, or other forms, of continuously declining discounting, $IRR^{T06} > IRR^{T012} > IRR^{T612}$.

Another, more recent version of the quasi-hyperbolic model involves a fixed, rather than variable, cost of receiving payments in the future (Benhabib et al., 2010). The fixed-cost specification allows this model to predict a "magnitude effect", a tendency for measured impatience to decrease as stake sizes increase, *ceteris paribus*. Aside from predicting a magnitude effect, however, this model has the same properties as the quasi-hyperbolic model: There is an extra cost of waiting, above and beyond exponential discounting, when comparing the present to the next future period; when comparing two adjacent, future periods, however, the extra cost applies to both periods, and thus cancels out, leaving discounting to be governed solely by the exponential discount rate. The quasi-hyperbolic

[21] The model is quasi-hyperbolic in the sense that it provides a step-function approximation to a hyperbolic discount function (see, e.g., Laibson, 1997).

model with fixed costs thus predicts $IRR^{T06} > IRR^{T012} > IRR^{T612}$ if the present extends further than two days into the future, or invariance of the IRR to time horizon, $IRR^{T06} = IRR^{T012} = IRR^{T612}$, if it does not, the same as the standard quasi-hyperbolic model.[22]

Relaxing the (usually implicit) assumption made by many previous studies, that people treat money rewards as consumption, changes the qualitative predictions for all non-constant discounting models. If payments are viewed as fully fungible, as predicted by standard theory, then subjects should use the experiment as an opportunity for arbitrage, and in all time horizons choose to wait when the annual rate of return exceeds the relevant market interest rate (this argument is made, e.g., by Coller and Williams, 1999, or Harrison et al., 2002). Outside of the experiment subjects can then use the credit market as appropriate, for example, borrowing against future experimental earnings using the lower market interest rate in order to finance their desire for present-biased consumption. All discounting models in this case predict $IRR^{T06} = IRR^{T012} = IRR^{T612}$, at the relevant interest rate faced by each individual. Thus, the ability of discounting models to explain the data assuming full arbitrage depends on whether IRRs are invariant to time horizon.

It has also been typical to assume that the unobserved utility function is linear, for the range of stakes used in discounting experiments, i.e., that earning 20 or 30 more Euros should not have an impact on the marginal utility of money. From some viewpoints, this assumption is plausible (see, e.g., Rabin, 2000), while another strand of the literature has

[22] In this model, a payment Z_t received in the future is discounted by $\Delta_t = \delta^t - \frac{b}{Z_t}$, where $\delta = \frac{1}{1+\rho}$ and $b > 0$ is a fixed cost of having a payment arrive in the future, which goes to zero as the stakes in the experiment increase. A payment received at $t = 0$ is not discounted, i.e., $\Delta_0 = 1$. The present value of a future payment is thus given by $\Delta_t Z_t = \delta^t Z_t - b$. Importantly, between-period discounting in the future is the same as in the exponential model (and as in the quasi-hyperbolic model with variable costs). I.e., the Euler equation for consumption in future periods $t+1$ and $t+2$ is $\delta^{t+1} Z_{t+1} - b = \delta^{t+2} Z_{t+2} - b \rightarrow Z_{t+1} = \delta Z_{t+2}$. Maintaining the assumption that the present extends more than two days into the future, the indifference conditions implied by choices are given by

$$(1+\frac{\rho}{2})100 + (1+\frac{\rho}{2})b = Z^{T06}; \qquad (1+\frac{\rho}{2})^2 100 + (1+\frac{\rho}{2})^2 b = Z^{T01212}; \qquad (1+\frac{\rho}{2})100 = Z^{T612}$$

Substituting into (2) yields

$$IRR^{T06} = 2((1+\frac{b}{100})(1+\frac{\rho}{2})-1); \qquad IRR^{T012} = 2((1+\frac{b}{100})^{\frac{1}{2}}(1+\frac{\rho}{2})-1); \qquad IRR^{T612} = 2((1+\frac{\rho}{2})-1)$$

These equations imply $IRR^{T06} > IRR^{T012} > IRR^{T612}$. If, instead, the present-bias falls within the window of two days from the present, all terms involving b are eliminated, the conditions reduce to those in (1), and the prediction is invariance of IRR with respect to time horizon, $IRR^{T06} = IRR^{T012} = IRR^{T612}$. Qualitative predictions are similar using the alternative specification of the model, discussed by Benhabib et al. (2010), where exponential discounting is applied to b as well as the future payment.

pointed to evidence of small-stakes risk aversion in lottery experiments, and interpreted this behavior as indicating that the marginal utility of income is diminishing even over small stakes (e.g., Andersen et al., 2008).[23] If utility is concave, then some of the apparent impact of time horizon length on IRRs could be due to non-linear utility. For example, imposing linearity could overestimate the true difference in IRRs between short and long time horizons, given the right type of concavity of the unobserved utility function.[24] As we discuss in detail later, however, even if utility is concave this is unlikely to be more than a partial explanation for the results, for quantitative reasons. Discrepancies in IRRs across time horizons are sufficiently large that it would require a highly implausible degree of concavity, far beyond typical utility calibrations based on lottery experiments or other data sources, to explain more than part of the difference.

If the assumptions about time separability and time stationarity of the utility function are not valid, then measured IRRs may vary freely across time horizons for reasons unrelated to non-constant discounting and dynamic inconsistency. For example, people may anticipate that the marginal utility of consumption in the future depends not just on distance from the present, due to discounting, but also on state-contingent preferences, for example an upcoming vacation. In this case, "anything goes" and the model is very flexible, so one could observe many different patterns of IRRs across time horizons, for any given discount function. Our approach is to assess whether the data can be explained by models that maintain the workhorse assumptions of separability and stationarity, which are used for most applications of discounted utility theory.

---

[23] There are alternative explanations for small-stakes risk aversion besides diminishing marginal utility, such as "loss aversion" (see, e.g., Rabin and Thaler, 2001).

[24] Allowing for concavity means that we overestimate the level of IRR for any given time horizon. If the estimates of impatience are affected by concavity of the unobserved utility function, we have $\widetilde{IRR}^{T06} = 2(\frac{u(Z^{T06})}{u(100)} - 1)$, $\widetilde{IRR}^{T012} = 2((\frac{u(Z^{T012})}{u(100)})^{\frac{1}{2}} - 1)$, and $\widetilde{IRR}^{T06} = 2(\frac{u(Z^{T612})}{u(100)} - 1)$. By concavity, it is clear that $\widetilde{IRR}^{T06} < IRR^{T06}$, $\widetilde{IRR}^{T012} < IRR^{T012}$, and $\widetilde{IRR}^{T612} < IRR^{T612}$. For our purposes, however, the more relevant question is how the *relative* sizes of IRRs for different time horizons change, allowing for concavity. There are two effects that may work in opposite directions. On the one hand, introducing concavity leads to a bigger reduction in the value of the ratio $Z^{Tst}/100$ for the longer time horizon, e.g., $Z^{T012}/100 - u(Z^{T012})/u(100) > Z^{T06}/100 - u(Z^{T06})/u(100)$. This works in the direction of reducing IRR^{T012} by more than the IRRs for short horizons, implying that we underestimate the true discrepancy between IRRs for short and long horizons by assuming linearity. On the other hand, a given change in the ratio in late payment to early payment causes a larger decrease in IRRs for short horizons than long horizons, due to compounding (this is immediately observable from the relevant derivatives of IRRs with respect to these ratios). Thus, whether or not concavity understates or exaggerates the difference in IRRs depends on the relative sizes of these different effects.

3 Results on IRR and time horizon

Figure 1 presents the cumulative distribution functions of the annual IRR for each of the different time horizons. The top panel shows results from the CSS data, and the bottom panel shows results from the SOEP data. For the SOEP data, we pool the T012 measures across the two sub-samples that have this treatment, as the order and stake sizes are identical.[25]

In the top panel of Figure 1, we see that people in the CSS data exhibit a different distribution of IRRs for T012 than for T06, or T612 ($p < 0.001$; $p < 0.001$; Kolmogorov-Smirnov), in the direction of lower IRRs and greater patience for the 12-month horizon. People are quite similar in impatience, however, comparing the distributions of IRRs for the T06 and T612 measures, and there is no significant difference ($p < 0.90$; Kolmogorov-Smirnov). Thus, people tend to be more patient for long horizons than short horizons, but similarly impatient over the short horizons regardless of the starting date. The bottom panel shows the same qualitative patterns in the SOEP data. Observed IRRs increase monotonically as time horizon length decreases, with greater impatience for T06 than T012 ($p < 0.001$; Kolmogorov-Smirnov), and even greater impatience for T01 than T06 ($p < 0.001$; Kolmogorov-Smirnov). At the same time, IRRs are relatively insensitive to the starting date of the time horizon, in that people are similarly impatient for T01b and T1213 ($p < 0.98$; Kolmogorov-Smirnov). The magnitudes of the differences in IRRs are also substantial. For example, in the CSS data, median IRRs for T06 and T612 are 10 percentage points higher than the median IRR for T012, and in the SOEP data, the IRR for T01 is more than 25 percentage points higher than the median IRR for T012.

The distribution for T01 is also similar to T01b and T1213, except for a deviation in the direction of greater patience starting around the middle of the range for T01, potentially due to either an order effect, or an effect of the different upper bound for the IRR, or both.[26] Despite this framing difference between the one-month measures, they are all clearly more similar to each other, than to measures with longer horizons.

---

[25] Empirically, the cumulatives for the two measures considered separately are very similar, and are not significantly different, using either parametric ($p < 0.849$; Kolmogorov-Smirnov) or non-parametric tests ($p < 0.539$; Mann-Whitney). Medians are also identical for the two sub-samples.

[26] The distributions for T01b and T1213 are each significantly different from the distribution for T01 ($p < 0.001$; $p < 0.001$; Kolmogorov-Smirnov).

Tables 2 and Table 3 provide another way to look at the results on IRR and time horizon, using interval regressions that correct for right- and left-censoring of the dependent variable. The dependent variable is the measured IRR, and independent variables are dummy variables for time horizon length. Standard errors are robust and adjusted to allow for potential correlation of the error term across observations from the same individual.

In Column (1) of Table 2 we see that IRRs in the CSS data are significantly lower for the T012 measure compared to T06, by more than 6 percentage points, while there is not a significant difference between T612 and T06. Table 3 presents similar regression analysis based on the SOEP data. Results are reported separately for the three sub-samples, in Panel A, B, and C, respectively. Looking at Column (1) we again see a pattern of lower IRRs for longer horizons, but similar IRRs across horizons of the same length regardless of the starting date: T012 is significantly lower than T06 by about 5 percentage points, and lower than T01 by about 20 percentage points, while T1213 is not significantly different from T01b and the point estimate indicates only a 1 percentage point difference.[27]

The regression analysis also allows checking robustness with respect to order effects, in the CSS data where order was randomized. We add dummy variables for the different possible treatment orders, and interactions of T012 and T612 with all of the different orders, to the specification used in Column (1) of Table 2. We again find a significant difference between T012 and T06, but not between T06 and T612. Furthermore, all interaction terms are not statistically significant, individually or jointly.[28] Thus, the qualitative results are systematic across all different possible treatment orders, and are not driven by any particular treatment ordering.[29]

In subsequent columns, Table 2 and Table 3 explore the robustness of the qualitative

[27] It is noteworthy that IRRs tend to be lower in the SOEP data than the CSS data, for comparable time horizons (e.g., compare the constant term in Column (1) of Table 2 to the constant term in Column (1), Panel A, of Table 3). One interpretation is a "magnitude effect", such that the level of impatience decreases with stake size for a given pair of horizons. Such an effect would be an anomaly for standard discounting models, where the IRR is assumed to be independent of stake size. Indeed, many studies report finding lower IRRs as stake sizes increase (see Frederick et al., 2002). An alternative explanation for this stylized fact, however, could be that the unobserved utility function is more linear for higher stakes (see Andersen et al., 2011, for a discussion).

[28] Results available upon request.

[29] We do observe that the impact of time horizon length is quantitatively smaller in first treatments, than in second or third treatments, i.e., the effect appears to be larger the more treatments individuals experience. The tendency for impatience to be greater in short than long horizons is still present even in first treatments, however, and the difference-in-differences for time horizon effects, comparing first treatment to second treatment, or first treatment to third treatment, or first treatment to later treatments, are not statistically significant.

results observed for the whole sample, across different sub-populations. We consider sub-populations defined by gender, age, cognitive ability, education level, income level, and credit constraints. Such variables might matter for how individuals respond to different measures of IRR, for a variety of reasons. For example, cognitive ability or education might be related to the level of financial sophistication of an individual, and familiarity with market interest rates, and potentially affect sensitivity to varying time horizon.[30] Income or credit constraints could potentially matter if people cannot borrow and have expenses that will arrive within the scope of the time horizons considered in the experiment.[31] Looking at the regressions estimates, it is apparent that the qualitative results on time horizon effects are quite robust, and do not appear to be driven by the degree of cognitive ability, financial sophistication, or financial situation: The difference between long and short time horizons is observed for every sub-sample, in both data sets, regardless of the measures used, while horizons of similar length, but different starting dates, are not significantly different.[32] Thus, our data indicate that the patterns we find are not isolated to specific populations, but rather are a quite robust and general tendency.

Similar conclusions arise from an individual-level analysis, showing that the aggregate results are not driven by any sort of misleading composition effect due to heterogeneous individual types. To identify individual "types", we focus on the CSS data, because there we have all three types of comparisons (OD, SD, and OSD) for each person and an ability to check whether the aggregate pattern is also present at the individual level. Table 4 reports fractions of individuals exhibiting different possible qualitative choice patterns across the different measures. Adhering to the usual maintained assumptions for testing discounting models, we classify individuals who consistently have the same IRR across all time horizons as constant discounters (this fraction could also include quasi-hyperbolic discounters, or subjects engaging in arbitrage, depending on assumptions). People with

[30] Cognitive ability is measured using two different tests, which are then standardized and averaged to provide a measure of overall cognitive ability. For a description of the tests see Dohmen et al. (2010).

[31] To create the dummy variable indicating being credit constrained, we use a question that asks: "If you suddenly encountered an unforeseen situation, and had to pay an expense of 1,000 Euros within the next two weeks, would it be possible for you to make that payment?"

[32] In previous work (Dohmen et al., 2010) we found a significant correlation between cognitive ability and impatience measured over an annual time horizon (T012). We replicate this finding in the CSS data, even when averaging across all three time horizons, and also in the SOEP for each sub-sample, averaging across time horizons. Thus, there is accumulating evidence that the *level* of IRR for an individual is related to cognitive ability. By contrast, the overall pattern of *differences* in IRR across time horizons that we observe are equally pronounced for low and high ability individuals.

$IRR_{T06} > IRR_{T012} > IRR_{T612}$ are classified as declining discounters, and people with the opposite pattern are classified as increasing. We exclude individuals who cannot be classified unambiguously, due to right-censoring of the IRRs.

One feature of the individual-level data that is immediately apparent, looking at the first row of Table 4, is that the modal pattern of individual choice matches the aggregate results; almost 50 percent of individuals are more impatient for both short time horizons than the longer time horizon, contrary to standard discounting models. Thus, the aggregate results are indicative of the most prevalent pattern of individual choices. The "Other" category is also inconsistent with discounting assumptions. It includes individuals who have lower IRRs in both short horizons than the long horizon, and those having at least one short horizon that is different from the long horizon, but the same IRR for the long horizon and the other short horizon. Overall, about 65 percent of individuals are inconsistent with all discounting types.

The first row of Table 4 also shows the frequencies of individuals who can be classified as being fully consistent with the predictions of constant, declining, or increasing discounters, which are 11.34, 11.63, and 10.76 percent, respectively. Thus, while there are some individuals who cannot be ruled out as being consistent with the usual candidate discounting models, these are relatively infrequent.[33]

Since the main message of the individual-level results is that most individuals make choices inconsistent with usual candidate discounting models, it is interesting to assess whether this is robust, in particular to potential "trembling hand" errors. One way to interpret time horizon sensitivity could be in terms of such (random) errors. To allow for relatively large trembling hand errors, we assess how the classifications change if we adopt an even weaker condition for equality of IRRs across time horizons. We convert the switching point for each time horizon into an interval 7.5 percentage points wide, by taking the upper and lower bounds and subtracting and adding 2.5 percentage points, respectively. We then classify IRRs as being different only if their respective intervals do

[33] Notably, there is a way in which the results shown in the first row of Table 4 may slightly overstate the frequency of individuals who are constant discounters. We have treated the IRRs for two time horizons as being equal if they have the same lower bound, but the true IRRs for each horizon could potentially be unequal, lying anywhere in the 2.5 percentage point interval starting from the lower bound. Thus, some of those classified as constant discounters could in fact be types who are inconsistent with discounting models, and some could be types who are consistent with declining or increasing discounting. This would mean that 65 percent being inconsistent with usual candidate discounting assumptions is a lower bound.

not overlap at all. This procedure is very conservative, in that true IRRs can differ by as much as 15 percentage points, and still be classified as equal. Clearly this procedure will tend to increase the frequency of individuals classified as constant discounters, for mechanical reasons, and systematically works against finding declining, increasing, or intransitive types. This is useful, however, as a hard test of the robustness of the main result of the individual level analysis: that a substantial fraction of individuals are not consistent with any of the discounting assumptions.

The second row of Table 4 shows the results. Not surprisingly, the fraction of constant discounters increases substantially, to 39.32 percent, and the fractions of declining and increasing discounters are smaller, 8.36 and 6.81 percent, respectively. The fraction of individuals who are inconsistent with standard discounting models remains substantial, however, and is still the relative majority among types, at approximately 46 percent of the sample. Thus, the conclusion that many individuals exhibit behavior that falls outside the bounds of the discounting framework is robust to allowing for even rather large random errors in decision making.

3.1 Discussion: Models and the stylized facts

To summarize, the empirical analysis yields two main stylized facts: (1) People are more impatient for short than long time horizons; (2) people are relatively insensitive to when a given time horizon starts. In the CSS data this is shown by the pattern $IRR^{T06} = IRR^{T612} > IRR^{T012}$, and in the SOEP data by $IRR^{T01b} = IRR^{T1213} \approx IRR^{T01} > IRR^{T06} > IRR^{T012}$.

In light of the predictions derived in Section 2.3, it is clear that this pattern is not well explained by either constant, declining, or increasing discounting, maintaining the usual identifying assumptions. The sensitivity of IRRs to time horizon length is inconsistent with constant discounting, while the insensitivity of IRRs to starting date of a given horizon is contrary to the key prediction of declining or increasing discounting (and also to recent two-system models that predict declining discounting; see, e.g., Fundenberg and Levine, 2012).

Importantly, this inconsistency is revealed by having a design with multiple types of time horizons comparisons. Focusing only on one type of comparison, the data would have seemed to strongly support one discounting assumption or another. In particular, the tendency to have higher IRRs in short time horizons than long horizons means that the data would seem to strongly conform to the declining discounting prediction, looking solely at OD comparisons. On the other hand, the relative insensitivity to starting date means that behavior is much closer to the prediction of constant discounting, looking only at SD comparisons. And finally, because of greater IRRs in short than long time horizons, independent of starting date, OSD type comparisons would seem to indicate a strong increasing discounting pattern. Thus, focusing on any one time horizon comparison could be misleading, by seeming to support a particular discounting assumption when in fact the data as a whole are not consistent with any of the usual candidate discounting assumptions.

Relaxing the assumptions that people treat money as consumption, the usual candidate models are still not consistent with the data. If people treat money as fully fungible, and therefore use the experiment as an opportunity for arbitrage, IRRs should be invariant to time horizon, contrary to the findings. The pattern is also inconsistent with non-constant discounting models where the discount rate is assumed to drop discretely at some point in the near future.[34] If the drop in the discount rate occurs two or more days into the future, then in our experimental design these models make the same predictions as those with continuously declining discounting, and the insensitivity of IRRs to start date is a puzzle. If the drop occurs before two days, then the model predicts constant IRRs across time horizons, because even the earliest payments are available starting two days in the future. In this case the finding that time horizon does strongly influence IRRs is a puzzle.

Allowing for concave utility over the stakes in the experiment also seems unlikely to explain the results. This is because the differences in IRRs across time horizons are too large to plausibly reflect concavity. For example, assuming a standard CRRA utility function, increasing the degree of concavity tends to reduce the size of IRR^{T06} and IRR^{T612} relative to IRR^{T012}, but one would need a coefficient of relative risk aversion, θ, equal to 35 to equalize the true IRRs for short and long horizons. This is the same magnitude as

[34] In addition to the quasi-hyperbolic model, and the fixed-cost version of that model by Benhabib et al. (2010), this class includes 'two-system" models that make similar predictions (Thaler and Shefrin, 1981; Fudenberg and Levine, 2006), and the fixed cost version of non-constant discounting discussed by Benhabib et al. (2010).

the coefficient value that would be needed to explain the equity premium puzzle, and is typically regarded as implausibly high (see Siegel and Thaler, 1997). Furthermore, recent empirical estimates of θ, from lottery experiments with similar stake sizes to our intertemporal choice experiments, are far smaller. The median parameter value is estimated to be in the range of 0.55 to 0.75 (Andersen et al., 2008; Andersen et al., 2010).[35] Using these parameter values, or even substantially larger values, the relative size of IRRs is largely unchanged. For example, moving from linear utility to utility with $\theta = 0.75$, the ratio of IRR^{T012} to IRR^{T06} for the median individual goes from 0.73 to 0.74, and the ratio of IRR^{T012} to IRR^{T612} goes from 0.69 to 0.70.[36] Other recent estimates, based on large-stakes (hypothetical) lottery questions administered in a large representative survey, show parameters in a similar range, and show that $\theta = 35$ is completely outside the support of the empirical distribution (see Dohmen et al., 2005). Thus, while the right type of unobserved concavity could in principal play a role in explaining time horizon effects, it seems unlikely that is explains more than a small portion of the effects that we observe.

We summarize the ability of different models to explain individual time horizon comparisons, as well as the data as a whole, in Table 5 (in the Appendix). The most important message of the table is that none of the models can accommodate the whole set of facts. The table is also potentially useful in that it suggests which types of designs may tend to find which types of patterns.

We also see in the CSS data, where there are three time horizon comparisons per individual, that the data are inconsistent with a broader class of models, because they imply a violation of transitivity. This is true focusing on the aggregate choice patterns, and also for the majority of subjects at the individual level, maintaining the usual assumptions of time stationarity and separability. Intuitively, the violation arises because the 12-month measure, and the two 6-month measures, give conflicting information about an individual's preferences, for instance in terms of the size of late payment needed to induce waiting 12

[35] In this approach, lottery experiments generate modest variations in earnings, and risk aversion is assumed to reflect diminishing marginal utility of money over this range.

[36] Following (2), the IRR ratios are calculated according to

$$\frac{\widetilde{IRR}^{T012}}{\widetilde{IRR}^{T06}} = \frac{\left(\frac{Z^{T012}}{100}\right)^{1-\theta} - 1}{\left(\frac{Z^{T06}}{100}\right)^{1-\theta} - 1}; \qquad \frac{\widetilde{IRR}^{T012}}{\widetilde{IRR}^{T06}} = \frac{\left(\frac{Z^{T012}}{100}\right)^{1-\theta} - 1}{\left(\frac{Z^{T612}}{100}\right)^{1-\theta} - 1} \tag{6}$$

using $\theta = 0.75$ and median late payments of 119, 129, and 120 Euros for T06, T012, and T612, respectively.

months. This is different from dynamic inconsistency, because the inconsistency is present at a given point in time.

More formally, note that T06, T012, and T612 pose individuals with choices between different bundles, where elements are payments received at 0, 6, or 12 months. In the data, we observe for each time horizon the late payment amounts Z^{T06}, Z^{T012}, and Z^{T612} such that an individual is indifferent between the bundle involving the early payment and the bundle involving the late payment. The table below shows these particular bundles for each horizon:

	$T06$		$T012$		$T612$
A	$(100, 0, 0)$	A	$(100, 0, 0)$	D	$(0, 100, 0)$
B	$(0, Z^{T06}, 0)$	C	$(0, 0, Z^{T012})$	E	$(0, 0, Z^{T612})$

where individuals have preferences $A \sim B$, $A \sim C$, and $D \sim E$ when presented with each choice pair in isolation. For now, we are assuming that utility is linear, and $u(x) = x$.

An inconsistency arises, however, when one compares across choice pairs. From T012, we observe the Z^{T012} that makes an individual indifferent between receiving 100 at time 0 or Z^{T012} in 12 months

$$Z^{T012} = \frac{100}{(1 + \frac{IRR^{T012}}{2})(1 + \frac{IRR^{T012}}{2})} = \frac{100}{(1 + \rho_1)(1 + \rho_2)}. \tag{7}$$

Combining the observed IRRs for T06 and T612, we can also construct a late payment $Z^{T012'}$ that should make the individual indifferent between 100 at time 0 or waiting 12 months to get $Z^{T012'}$

$$Z^{T012'} = \frac{100}{(1 + \frac{IRR^{T06}}{2})(1 + \frac{IRR^{T612}}{2})} = \frac{100}{(1 + \tilde{\rho}_1)(1 + \tilde{\rho}_2)}. \tag{8}$$

For the typical individual, however, the differences in IRRs that we observe for short and long time horizons imply

$$\left(\frac{1}{1 + \rho_1}\right)\left(\frac{1}{1 + \rho_2}\right) \neq \left(\frac{1}{1 + \tilde{\rho}_1}\right)\left(\frac{1}{1 + \tilde{\rho}_2}\right). \tag{9}$$

I.e., total discounting over 12 months is different, based on the 6-month measures, than it is based on the 12-month measure. This immediately implies (from (7) and (8)) that $Z^{T012} \neq Z^{T012'}$.

To see how this inconsistency translates into a violation of transitivity, note that the 6-month measures say an individual should have preference $A \sim C'$ if posed with a choice

$$\begin{array}{ll} & T012' \\ A & (100, 0, 0) \\ C' & (0, 0, Z^{T012'}) \end{array}$$

and since $Z^{T012} \neq Z^{T012'}$, should not be indifferent for Z^{T012}. I.e., $A \sim C' \implies A \nsim C$. But combining these implied preferences, with the preferences that are actually revealed in T012 for late payment Z^{T012}, we have $A \sim C \nsim C' \sim A$. This implies $A \sim C \nsim A$, a violation of transitivity.

There is also a violation of transitivity relaxing the assumption of linear utility, and allowing for a concave (monotonic) utility function $u(\cdot)$, as long as (9) holds. As discussed above, (9) holds even if utility is concave, unless the utility function is characterized by implausibly extreme concavity. With non-linear utility the late payments satisfy the following conditions:

$$u(Z^{T012}) = \frac{u(100)}{(1+\rho_1)(1+\rho_2)}; \qquad u(Z^{T012'}) = \frac{u(100)}{(1+\tilde{\rho}_1)(1+\tilde{\rho}_2)}. \tag{10}$$

.

Given $(1+\rho_1)(1+\rho_2) \neq (1+\tilde{\rho}_1)(1+\tilde{\rho}_2)$ we clearly have $Z^{T012} \neq Z^{T012'}$, and there is an inconsistency in preferences that again implies a violation of transitivity.

If one relaxes the assumptions of time stationarity of utility, and separability, the data need not imply a violation of transitivity. If the utility function can vary arbitrarily over time, or the discount factor enters the period utility function, then "anything goes". However, these are usual workhorse assumptions of the discounted utility framework, and so a violation conditional on these assumptions is still relevant for the types of models used in a wide variety of applications. Referring again to Table 4, all individuals except those who fall into the strictly constant, declining, or increasing categories have choice patterns that imply violations of transitivity, maintaining the assumptions of separability at time stationarity. This is true for *any* arbitrary discounting function.

4 Discussion

The results in this paper pose an important puzzle for understanding inter-temporal choice. People are more impatient for short than long time horizons, contrary to constant discounting, but insensitive to starting date, in contrast to non-constant discounting models where the starting date is crucial. This pattern of inter-temporal choice is also inconsistent with a broader class of models, in that it implies a violation of transitivity. The phenomenon is robust to real incentives, it is pervasive in the population as a whole as well as in different sub-groups, and it replicates across different data sets and different types of measures. These results have implications for interpreting previous empirical literatures, and suggest new directions for modeling inter-temporal choice.

4.1 Experimental Measures and Life Outcomes

One implication of our results concerns how to interpret the previous empirical literature on time horizon effects and self-control problems. Previous studies (see footnote 2) have had only limited success in predicting life outcomes thought to reflect self-control problems (smoking, high body mass index, overspending, etc.), using experimental measures of time horizon effects. One possible explanation, suggested by our findings, is that some other underlying mechanism besides non-constant discounting causes "time horizon sensitivity". Indeed, time horizon effects turn out to be primarily inconsistent with non-constant discounting models. This would help explain why time horizon effects are not that strongly related to life outcomes chosen specifically because of their self-control problem character. Our results also suggest another explanation, however, that would apply even if some time horizon effects do reflect declining or increasing discounters. To date, empirical analyses have typically been based on a single time horizon comparison, and thus those who are candidates for being, say, a true constant discounter, are potentially lumped together with many types who are actually intransitive but are misclassified. In this case, attenuation bias could contribute to the weak results.

To shed some light on whether indicators of declining or increasing discounting are strong predictors of self-control problems in life, once one corrects for mis-classification and reduces attenuation bias, we present a final set of results from our CSS data. In the CSS data we have three measures for each individual, which makes it possible to separately identify constant, declining, increasing, and intransitive choice patterns. Table 6 (in the Appendix) reports regression results, where we look at a range of life outcomes: smoking, poor nutrition, drinking, body mass index, days per year with an overdrawn checking account, log net income, self-reported tendency to save, and an overall index of these outcomes. These types of outcomes have been examined in previous studies, because of their presumed links to self-control problems, although typically a given study has considered only a few rather than the whole set. As a control variable, we include the average IRR for an individual across all three time horizons, to account for the possibility that different time horizon patterns might be correlated with the overall level of IRR. We also include a set of controls for personal characteristics, including gender, age, cognitive ability, and education, which are similar to those used by Chabris et al. (2008).[37] To our knowledge, ours is the first empirical analysis on how measures of time horizon effects relate to life outcomes, using a specification that identifies intransitive types.

Table 6 shows that indicators for declining or increasing discounting are only modestly successful at predicting self-control problems in life, even in a specification that helps address attenuation bias.[38] In unreported regressions, we also used a specification without an indicator for intransitive types, so that these were included in the omitted category with constant discounters. In that case we found that indicators for increasing discounting and declining discounting patterns were even less strongly related to outcomes; the only significant result was for declining and the overall index. Thus, including the intransitivity indicator does sharpen predictions, consistent with reducing attenuation bias, but the resulting predictive success is still modest. It is also relevant that, in Table 6, the relationships of increasing and declining choice patterns in the experiment to outcomes in life are qualitatively the same. This is in contrast to what one might expect if these capture different shapes of discount functions; declining and increasing discounters should have opposite forms of self-control problems, such that increasing discounters are, e.g., relatively less likely to smoke, or overspend, than constant or declining discounters. Thus,

[37] Results are less precisely estimated, but qualitatively similar, without controls for personal characteristics. Intransitivity, declining, and increasing are all significantly related to one or two outcomes, and the average IRR is significantly related to more outcomes.

[38] Results are based on excluding individuals who have censored IRRs for any of the time horizons. We have similar results if we also include individuals who have just one censored IRR; these are unambiguously classified in terms of constant, declining, increasing, or intransitive, type, but the there is some imprecision in that the average level of the IRR is calculated using one right-censored value.

in terms of predictive success, and qualitative results, these findings are more consistent with the first explanation: that time horizon effects in experiments may reflect some other mechanism besides non-constant discounting and self-control problems.

We did also analyze the relationship between time horizon effects and life outcomes using the SOEP data. There we have the same types of life outcomes as in the CSS, and typically about 1,000 individuals for our estimation sample (those with non-censored observations, and non-missing information for control variables and a given outcome). In this case, however, it is not possible to identify intransitive types, because of having only one type of time horizon comparison per individual. We find that indicators for declining and increasing discounting, constructed appropriately for each sub-sample, are related to very few life outcomes. The only significant correlation is between declining discounting and the self-reported indicator for being a saver. With the caveat that attenuation bias is more of a problem in this case, we again find a weak correlation between time horizon effects and self-control-problem outcomes. This is as one would expect, if time horizon sensitivity in experiments largely reflects some other mechanism besides self-control problems.

Notably, the best all-around predictor of life outcomes in Table 6 is the average level of IRR, and the correlations are all in the directions one would expect, based on a tendency to weight future consumption less heavily than the present.[39] In the SOEP, we also find that the level of IRR is a good predictor of various life outcomes (see Vischer et al., 2011). Thus, the life outcomes we consider do appear to have an important intertemporal choice component, and the experimental approach does provide a useful way to predict such behaviors, even if time horizon effects in the experiments are only weakly related to this set of outcomes.

4.2 Concluding Remarks and Directions for Future Research

In terms of directions for future research, our results highlight the potential value of developing new types of experimental designs for measuring self-control problems.[40] One

[39] Results for IRR are similar and robust to dropping the indicator for intransitivity, changing definitions of censored individuals, and excluding controls for demographics. The results are quite consistent with those of, e.g., Chabris et al. (2008), who also use the level of discounting (instantaneous discount rate at time zero) and find similar relationships to a similar set of outcomes, or to Harrison et al. (2002).

[40] Noor (2011) proposes an interesting, alternative approach to measuring time discounting with monetary payments. Our results are complementary, by suggesting the potential value of incorporating multiple types of time horizon comparison.

possibility is to study whether behavior is more compatible with non-constant discounting models if one uses actual consumption opportunities, rather than monetary payments. Some studies have found patterns consistent with declining discounting in an SD design using food rewards over the course of minutes (McClure et al., 2007) or weeks (Read and van Leeuwen, 1998; Read et al., 1999). It has not yet been shown, however, whether a consumption-based experiment with a comprehensive design in terms of multiple types of time horizon comparisons, yields the same patterns of intransitivity as we find. If the goal is to capture self-control problems, other approaches might be also useful, for example building on on research showing that willingness to pay increases with degree of salience, in terms of whether or not a good is physically present (see, e.g., Bushong et al., 2010). Individual differences in sensitivity to such cues of immediate availability might prove to be a good predictor of self-control problems in life.

In terms of future work on modeling inter-temporal choice, our results suggest some potentially useful directions. One approach could be to maintain the usual structure of discounting models, except for the assumption that discounting is indexed to calendar date. The model could assume that people discount late payments relative to early payments, and do so in a way that makes the IRR lower for longer durations between early and late payments (e.g., with discount rates that decline with distance from early to late payment). Crucially, however, the model would assume discounting is independent of the calendar dates of early and late payments.[41] Consistent with our results, such a model would imply that people need to be compensated for waiting, but in a similar way regardless of when the waiting should occur.

An alternative type of approach, which involves a more radical departure from discounting models, has been proposed in various forms by Leland (2002), Rubinstein (2003), and Scholten and Read (2010). These models assume that individuals make inter-temporal choices through some kind of algorithm based on similarity judgements. For example, in Rubinstein (2003), individuals use two similarity relations, one for time duration and one for monetary amounts, and go through a three step process: (1) look for strict dominance, in the sense that one option dominates along both time and money dimensions, and choose

[41] See Scholten and Read (2006) and Ebert and Prelec (2007) for related discussions, on duration-dependent discounting and time-sensitivity in discounting, respectively. Another, interesting approach to modeling inter-temporal choice, in terms of attentional focus, has been proposed by Koszegi and Szeidl (2011). While the model generates many novel predictions about inter-temporal decision making, it does not explain the main pattern of behavior observed in our experiments.

this strictly dominant option if it exists; (2) if strict dominance fails, but options are judged similar on one dimension, then decide based on the dissimilar dimension; (3) if options are similar along all dimensions, decide "in some other way", potentially at random. In the context of our experiments, if people were to view differences in monetary amounts across long and short time horizons as significant, but differences in time horizon length, or starting date, as less important, this would generate greater patience for long horizons, where late payments are larger. It would also generate similar impatience across horizons that differ only in terms of timing but not monetary amounts. More work is needed, however, to assess whether people indeed follow such a procedure in making inter-temporal choices, and on understanding how similarity judgments are made.

A third approach would be to maintain canonical assumptions about time discounting, but allow for biased perceptions of interest rates. In particular, there is evidence that people are prone to a particular bias, "exponential growth bias", that involves ignoring the role of compounding (e.g., Eisenstein and Hoch, 2005). In the context of inter-temporal choice experiments, this type of bias could mean that individuals associate a given late payment with a rate of return that is higher than the true rate of return, because they fail to account for compounding. For example, in T012, neglecting the semi-annual compounding would cause an individual to overestimate the rate of return implied by a given late payment. The same individual would be un-biased, however, in their perception of the interest rates for the semi-annual time horizons T06, and T612. In combination with constant time discounting, this bias would generate a pattern of greater impatience in short than long time horizons, but constant discounting for horizons of the same length, irrespective of start date. However, while this particular bias may play a role, it is unlikely to be the main driver of our results, because the time horizons we consider are too short for compounding to make a large difference. For example, for the median late payment for T012 in the CSS data, ignoring compounding leads to overestimating the annual rate of return by 2 percentage points, whereas the median difference in IRRs for 12-month versus 6-month horizons 10 percentage points.[42] Nevertheless, investigating the role of biased perceptions more generally in generating time horizon effects is a potentially fruit-

[42] Some other pieces of evidence that argue against this bias explanation are the fact that the same patterns are observed for low and high cognitive ability individuals, and more and less educated individuals. Previous studies, by contrast, have found a strong positive correlation between lower educational attainment and biased perceptions of interest rates (Stango and Zinman, 2009).

ful direction for future research (see also Zauberman et al., 2009, and Ebert and Prelec, 2007).

Finally, the method variance we identify is potentially relevant for policy. There is an on-going discussion about interventions designed to address self-control problems in the form of over-spending or over-consumption of harmful goods (e.g., Camerer et al., 2003). Such policies are often motivated with reference to declining discounting. While there are many good reasons to think that self-control problems exist,[43] the impact of such policies may depend crucially on the types of self-control problems present in the population, and on the extent of heterogeneity. The current state-of-the-art for identifying the distribution of different types of self-control problems in the population, however, is to use monetary choice experiments. Our results show that the particular comparison used can matter strongly and systematically for the proportions of different types. For example, OD comparisons in our CSS data imply roughly 70 percent declining discounters, while OSD comparisons imply 70 percent increasing discounters (for the same subjects). Furthermore, the number of individuals who are consistent with any type of discounting model at all may in fact be quite small. More fundamentally, it may not make sense to equate time horizon effects with self-control problems if these reflect a different underlying mechanism. Therefore, our results show that developing alternative measures of self-control problems may also be important from a policy perspective.

References

Anderhub, V., W. Güth, U. Gneezy, and D. Sonsino (2001): "On the interaction of risk and time preferences: An experimental study," *German Economic Review*, 2(3), 239–253.

Andersen, S., G. Harrison, M. Lau, and E. Rutstroem (2010): "Discounting behavior: A reconsideration," *Georgia State Economics working paper.*

——— (2011): "Discounting Behavior and the Magnitude Effect," *Georgia State Economics working paper.*

Andreoni, J., and C. Sprenger (2010): "Estimating time preferences from convex budgets," *NBER Working Paper No. 16347.*

[43] Various pieces of evidence, outside of the context of monetary choice experiments, point to the existence of self-control problems. These include use of commitment devices, for example in signing up for gym contracts, and various instances of preference reversals in consumption data. See, e.g., Read and Van Leeuwen (1998), Della Vigna and Malmendier (2006); Ariely and Wertenbroch (2002); Ashraf et al. (2006); Milkman et al. (2009).

Ariely, D., and K. Wertenbroch (2002): "Procrastination, deadlines, and performance: Self-control by precommitment," *Psychological Science*, 13(3), 219–224.

Ashraf, N., D. Karlan, and W. Yin (2006): "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines," *Quarterly Journal of Economics*, 121(2), 635–672.

Baron, J. (2000): "Can we use human judgments to determine the discount rate?," *Risk Analysis*, 20(6), 861–868.

Bauer, M., J. Chytilova, and J. Morduch (2010): "Behavioral foundations of microcredit: Experimental and survey evidence from rural India," *IZA Discussion Paper No. 4901*.

Benhabib, J., A. Bisin, and A. Schotter (2010): "Present-bias, quasi-hyperbolic discounting, and fixed costs," *Games and Economic Behavior*, 69(2), 205–223.

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

Burks, S., J. Carpenter, L. Goette, and A. Rustichini (2011): "Which Measures of Time Preference Best Predict Outcomes? Evidence from a Large-Scale Field Experiment," *IZA Discussion Paper No. 5808*.

Bushong, B., L. King, C. Camerer, and A. Rangel (2010): "Pavlovian processes in consumer choice: The physical presence of a good increases willingness-to-pay," *American Economic Review*, 100(4), 1556–1571.

Camerer, C., S. Issacharoff, G. Loewenstein, T. O'Donoghue, and M. Rabin (2003): "Regulation for Conservatives: Behavioral Economics and the Case for" Asymmetric Paternalism"," *University of Pennsylvania Law Review*, 151(3), 1211–1254.

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

Coller, M., and M. Williams (1999): "Eliciting individual discount rates," *Experimental Economics*, 2(2), 107–127.

DellaVigna, S., and U. Malmendier (2006): "Paying not to go to the gym," *American Economic Review*, 96(3), 694–719.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2010): "Are Risk Aversion and Impatience Related to Cognitive Ability?," *American Economic Review*, 100(3), 1238–1260.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2005): "Individual Risk Attitudes: New Evidence from a Large, Representative, Experimentally-Validated Survey," *IZA Discussion Paper No. 1730*.

Ebert, J., and D. Prelec (2007): "The fragility of time: Time-insensitivity and valuation of the near and far future," *Management Science*, 53(9), 1423–1438.

Eisenstein, E., and S. Hoch (2005): "Intuitive compounding: framing, temporal perspective, and expertise," *Temple University working paper.*

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

Fudenberg, D., and D. Levine (2006): "A Dual-Self Model of Impulse Control," *American Economic Review*, 96(5), 1449–1476.

——— (2012): "Timing and Self-Control," *Econometrica*, 80(1), 1–42.

Halevy, Y. (2011): "Time Consistency: Stationarity and Time Invariance," *mimeo, University of British Columbia.*

Harrison, G., M. Lau, and E. Rutstroem (2005): "Dynamic Consistency in Denmark: A Longitudinal Field Experiment," *UCF Economics working paper.*

——— (2010): "Individual Discount Rates and Smoking: Evidence from a Field Experiment in Denmark," *Journal of Health Economics*, 29(5), 708–717.

Harrison, G. W., M. I. Lau, and M. B. Williams (2002): "Estimating Individual Discount Rates in Denmark: A Field Experiment," *American Economic Review*, 92(5), 1606–1617.

Kirby, K., and R. Herrnstein (1995): "Preference reversals due to myopic discounting of delayed reward," *Psychological Science*, 6(2), 83–89.

Koszegi, B., and A. Szeidl (2011): "A model of focusing in economic choice," *University of California, Berkeley, Economics Working Paper.*

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

Loewenstein, G., and D. Prelec (1992): "Anomalies in intertemporal choice: Evidence and an interpretation," *The Quarterly Journal of Economics*, 107(2), 573–597.

McClure, S., K. Ericson, D. Laibson, G. Loewenstein, and J. Cohen (2007): "Time discounting for primary rewards," *The Journal of Neuroscience*, 27(21), 5796–5804.

Meier, S., and C. Sprenger (2008): "Discounting financial literacy: time preferences and participation in financial education programs," *IZA Discussion Paper No. 3507.*

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

Milkman, K., T. Rogers, and M. Bazerman (2009): "Highbrow films gather dust: Time-inconsistent preferences and online DVD rentals," *Management Science*, 55(6), 1047–1059.

Noor, J. (2011): "Time Preference: Experiments and Foundations," *mimeo Boston University.*

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

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

Rabin, M. (2000): "Risk Aversion and Expected Utility Theory: A Calibration Theorem," *Econometrica*, 68(5), 1281–1292.

Rabin, M., and R. Thaler (2001): "Anomalies: Risk Aversion," *The Journal of Economic Perspectives*, 15(1), 219–232.

Read, D. (2001): "Is time-discounting hyperbolic or subadditive?," *Journal of Risk and Uncertainty*, 23(1), 5–32.

Read, D., M. Airoldi, and G. Loewe (2005): "Intertemporal tradeoffs priced in interest rates and amounts: A study of method variance," *working paper, London School of Economics and Political Science*.

Read, D., G. Loewenstein, and S. Kalyanaraman (1999): "Mixing virtue and vice: Combining the immediacy effect and the diversification heuristic," *Journal of Behavioral Decision Making*, 12(4), 257–273.

Read, D., and P. Roelofsma (2003): "Subadditive versus hyperbolic discounting: A comparison of choice and matching," *Organizational Behavior and Human Decision Processes*, 91(2), 140–153.

Read, D., and B. Van Leeuwen (1998): "Predicting hunger: The effects of appetite and delay on choice," *Organizational Behavior and Human Decision Processes*, 76(2), 189–205.

Roelofsma, P., and D. Read (2000): "Intransitive intertemporal choice," *Journal of Behavioral Decision Making*, 13(2), 161–177.

Rubinstein, A. (2003): "Economics and Psychology? The Case of Hyperbolic Discounting," *International Economic Review*, 44(4), 1207–1216.

Schupp, J., and G. G. Wagner (2002): "Maintenance of and Innovation in Long-Term Panel Studies The Case of the German Socio-Economic Panel (GSOEP)," *Allgemeines Statistisches Archiv*, 86(2), 163–175.

Siegel, J., and R. Thaler (1997): "Anomalies: the equity premium puzzle," *The Journal of Economic Perspectives*, 11(1), 191–200.

Stango, V., and J. Zinman (2009): "Exponential growth bias and household finance," *The Journal of Finance*, 64(6), 2807–2849.

Sutter, M., M. Kocher, D. Rutzler, and S. Trautmann (2010): "Impatience and uncertainty: Experimental decisions predict adolescents field behavior," *IZA Discussion Paper No. 5404*.

Takeuchi, K. (2010): "Non-parametric test of time consistency: Present bias and future bias," *Games and Economic Behavior*, 71(2), 456–478.

Thaler, R., and H. Shefrin (1981): "An economic theory of self-control," *The Journal of Political Economy*, 89(2), 392–406.

Thompson, S. K. (2006): "Targeted Random Walk Designs," *Survey Methodology*, 32(1), 11–24.

Vischer, T., T. Dohmen, A. Falk, D. Huffman, J. Schupp, U. Sunde, and G. Wagner (2011): "Measuring Time Preferences in Representative Samples by Combining Experimental Elicitation and Survey Measures," *University of Bonn Working Paper*.

Wagner, G. G., J. R. Frick, and J. Schupp (2007): "Enhancing the Power of Household Panel Studies: The Case of the German Socio-Economic Panel Study," *Schmollers Jahrbuch, forthcoming.*

Zauberman, G., B. Kim, S. Malkoc, and J. Bettman (2009): "Discounting time and time discounting: Subjective time perception and intertemporal preferences," *Journal of Marketing Research*, 46(4), 543–556.

Figures and Tables

Figure 1: Cumulative Distributions of IRR

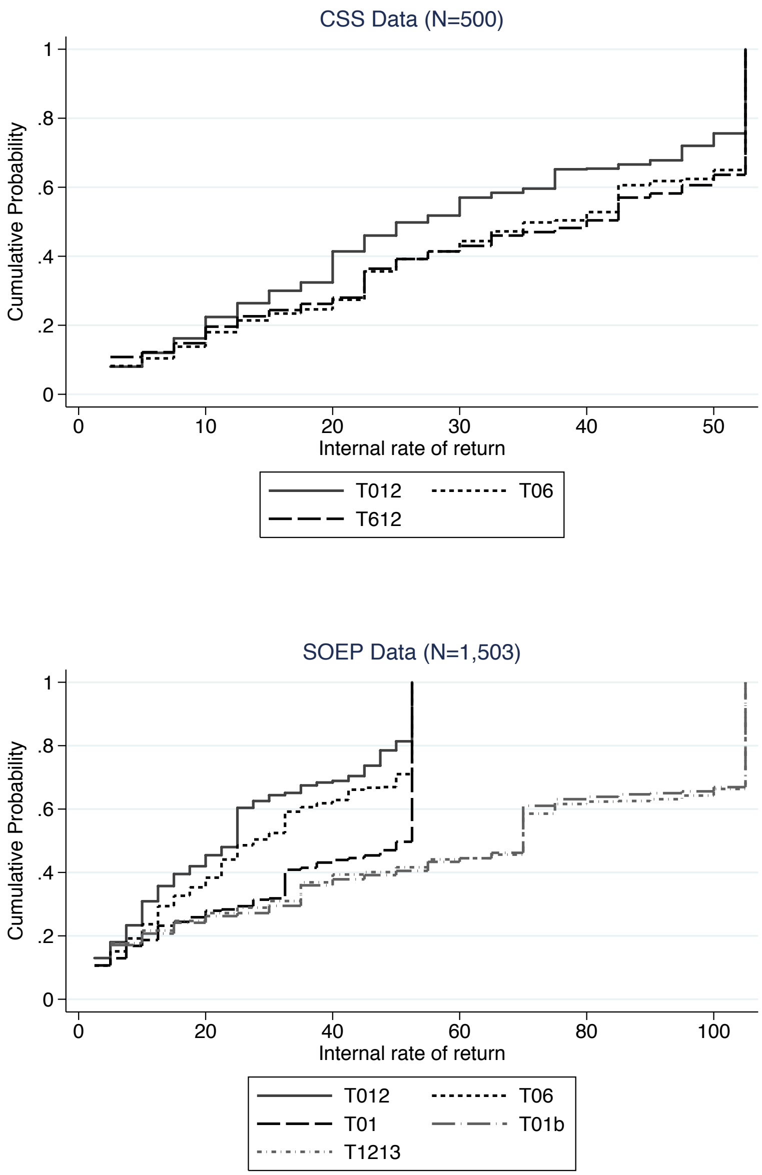

Table 1: Summary of Treatments

Measure	Data set	Sub-sample	Early payment (in Euro)	Upper-bound IRR	Obs.
T012	CSS	n.a.	100	52.5%	500
T06	CSS	n.a.	100	52.5%	500
T612	CSS	n.a.	100	52.5%	500
T012	SOEP	1 & 2	200	52.5%	977
T06	SOEP	1	200	52.5%	490
T01	SOEP	2	200	52.5%	487
T01b	SOEP	3	200	105%	526
T1213	SOEP	3	200	105%	526

Table 2: IRR as a Function of Time Horizon, by Demographic Groups: CSS Data

	All (1)	Males (2)	Females (3)	Age ≤med. (4)	Age >med. (5)	IQ ≤med. (6)	IQ >med. (7)	Less educated (8)	More educated (9)	Income ≤med. (10)	Income >med. (11)	Not credit Constrained (12)	Credit Constrained (13)
T12	-6.42***	-8.20***	-4.82***	-6.56***	-6.34***	-6.09***	-6.66***	-6.63***	-5.83***	-5.75***	-6.85***	-6.64***	-5.03**
	(0.85)	(1.20)	(1.19)	(1.05)	(1.39)	(1.53)	(0.96)	(1.04)	(1.42)	(1.32)	(1.10)	(0.92)	(2.17)
T612	0.14	0.25	0.07	-1.26	1.75	2.22	-1.38	0.71	-1.30	2.05	-1.11	-0.77	3.80
	(1.11)	(1.50)	(1.63)	(1.47)	(1.70)	(1.68)	(1.48)	(1.32)	(2.08)	(1.72)	(1.45)	(1.21)	(2.91)
Constant	37.22***	35.96***	38.33***	35.76***	39.14***	39.01***	35.95***	39.21***	31.84***	39.08***	35.90***	34.97***	42.32***
	(1.40)	(1.98)	(1.99)	(1.71)	(2.34)	(2.29)	(1.77)	(1.69)	(2.45)	(2.08)	(1.87)	(1.56)	(3.11)
Observations	1500	693	807	768	732	657	843	1113	387	591	909	1131	330

Notes: Interval regression estimates. Dependent variable is the IRR for a given time horizon. The reference time horizon is T06. Age, IQ, and household income groups are defined relative to median values. More educated indicates that an individual completed the Abitur, a college entrance exam in Germany. Credit constraints are measured by a question asking about ability to borrow money in the event of an unexpected expense. In parentheses, robust s.e., adjusted for clustering on individual. *, ** indicates significance at 10 and 5 percent level.

Table 3: IRR as a Function of Time Horizon, by Demographic Groups: SOEP Data

	All	Males	Females	Age ≤med.	Age >med.	IQ ≤med.	IQ >med.	Less educated	More educated	Income ≤med.	Income >med.	Not credit Constrained	Credit Constrained
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Panel A													
Sample 1: T012 vs. T06													
T012	-5.26***	-4.86***	-5.66***	-5.90***	-4.58***	-3.08***	-6.55***	-5.33***	-4.97***	-3.20**	-5.77***	-5.33***	-5.07***
	(0.68)	(0.89)	(1.03)	(0.94)	(0.98)	(0.93)	(0.93)	(0.77)	(1.44)	(1.49)	(0.76)	(0.72)	(1.87)
Constant	31.17***	28.31***	34.15***	30.48***	31.85***	31.87***	30.74***	34.11***	21.46***	29.91***	31.49***	27.63***	43.85***
	(1.34)	(1.86)	(1.91)	(1.87)	(1.91)	(2.13)	(1.72)	(1.62)	(2.17)	(2.98)	(1.49)	(1.41)	(3.22)
Observations	980	504	476	486	494	358	622	730	218	198	782	744	230
Panel B													
Sample 2: T012 vs. T01													
T012	-19.67***	-21.05***	-18.46***	-19.10***	-20.28***	-18.63***	-20.26***	-20.15***	-16.64***	-19.90***	-19.60***	-17.87***	-27.41***
	(1.35)	(2.04)	(1.80)	(1.97)	(1.86)	(2.12)	(1.75)	(1.62)	(2.39)	(2.64)	(1.57)	(1.44)	(3.70)
Constant	44.96***	46.61***	43.51***	41.08***	48.84***	46.24***	44.17***	47.50***	32.35***	42.77***	45.80***	41.10***	60.49***
	(1.80)	(2.64)	(2.47)	(2.42)	(2.70)	(2.83)	(2.34)	(2.11)	(3.51)	(3.18)	(2.18)	(1.94)	(4.49)
Observations	974	444	530	468	506	352	622	766	180	238	736	762	212
Panel C													
Sample 3: T1213 vs. T01													
T1213	-1.45	-0.22	-2.42	-0.10	-3.14	2.32	-3.15	1.78	-7.48	-6.56	0.23	-0.93	-3.60
	(3.36)	(5.03)	(4.52)	(4.33)	(5.27)	(6.28)	(3.95)	(4.15)	(6.03)	(6.94)	(3.84)	(3.70)	(7.12)
Constant	97.25***	96.53***	97.83***	91.92***	103.49***	118.21***	85.41***	102.48***	82.34***	96.81***	97.37***	89.92***	139.69***
	(5.16)	(7.72)	(6.94)	(6.79)	(7.89)	(9.91)	(5.86)	(6.36)	(8.85)	(10.48)	(5.93)	(5.25)	(18.21)
Observations	1052	464	588	552	500	424	628	784	240	254	798	860	188

Notes: Interval regression estimates, separately by sub-sample of the SOEP data. Dependent variable is the IRR for a given time horizon, with two time horizons (observations) per individual. Reference category is T06 in Panel A, T01 in Panel B, and T01b in Panel C. Age, IQ, and household income groups are defined relative to median values. More educated indicates that an individual completed the Abitur, a college entrance exam in Germany. Credit constraints are measured by a question asking about ability to borrow money in the event of an unexpected expense. In parentheses, robust s.e., adjusted for clustering on individual. *, ** indicates significance at 10 and 5 percent level.

Table 4: Individual types, CSS data

Types	Constant	Declining	Increasing	$IRR^{T06} = IRR^{T612}$ & $IRR^{T612} > IRR^{T012}$	$IRR^{T06} \neq IRR^{T612}$ & $IRR^{T06} > IRR^{T012}$ & $IRR^{T612} > IRR^{T012}$	Other
Percent	11.34	11.63	10.76	20.06	25.58	20.64
Types allowing for "error"	Constant	Declining	Increasing	$IRR^{T06} = IRR^{T612}$ & $IRR^{T612} > IRR^{T012}$	$IRR^{T06} \neq IRR^{T612}$ & $IRR^{T06} > IRR^{T012}$ & $IRR^{T612} > IRR^{T012}$	Other
Percent	39.32	8.36	6.81	17.65	4.64	23.22

Notes: In the first row, the sample (N=344) excludes individuals for whom right-censoring prevents unambiguous classification: Individuals with censoring in two or more time horizons. The category "Other" includes individuals who violate discounting predictions in the various other ways: IRRs for both short horizons are less than for the long horizon; IRR for one short horizon is greater (less) than IRR for the longer horizon, while IRR for the other short horizon is equal to IRR for the long horizon. In the second row, we allow for errors: for each horizon, we construct an interval by subtracting and adding 2.5 percentage points to the lower and upper bound IRRs, respectively, for each time horizon. We count two horizons as being equal if their intervals overlap. The same restrictions apply, in terms of excluding censored observations, but 21 additional individuals are excluded (N=323). Of the excluded individuals, 8 exhibit "cycling" or intransitivity in the relations between the three horizon measures, due to the ways that intervals do or do not overlap. E.g., $IRR_{06} = IRR_{012}$, $IRR_{06} > IRR_{612}$, but $IRR_{012} = IRR_{612}$. The remaining 13 that are excluded are individuals who become unclassifiable, because one of the non-censored IRRs now overlaps a right-censored IRR.

A Appendix (not for publication)

A.1 Additional Tables

Table 5: Stylized Facts on IRR and Time Horizon, and Different Models

	Constant discounting	Declining discounting	Increasing discounting	$\beta-\delta$ or similar, present$\leq$ 2 days	Discounting with arbitrage
CSS data					
Finding 1					
T06>T012		Yes			
Finding 2					
T06=T612	Yes			Yes	Yes
Finding 3					
T612>T012			Yes		
SOEP data					
Finding 4					
T06>T012		Yes			
Finding 5					
T01>T012		Yes			
Finding 6					
T01>T06		Yes			
Finding 7					
T01b=T1213	Yes			Yes	Yes

Notes: Findings compare mean (median) IRRs for different time horizons. Table entries of "Yes" indicate stylized facts that a given model can explain. Declining discounting includes, but is not limited to, hyperbolic discounting, quasi-hyperbolic discounting with the present > 2 days, and fixed cost discounting assuming present > 2 days. $\beta-\delta$ or similar, with present $\leq$ 2 days, includes quasi-hyperbolic discounting, fixed cost quasi-hyperbolic discounting, and two-system models that also have a discrete drop in the discount rate. Discounting with arbitrage refers to predictions, for any shape of discount function, when individuals do not treat monetary payments as consumption.

Table 6: Relating experimental measures to outcomes: CSS data

	Increasing	Declining	Intransitive	Avg. IRR	Obs.
BMI	1.02	1.39	0.21	-0.42	243
	[0.90]	[1.23]	[0.65]	[0.36]	
Smoker	0.24**	0.25**	0.17**	0.11***	249
	[0.12]	[0.11]	[0.06]	[0.04]	
Drinking	0	0	0.06	0.05	249
	[0.21]	[0.21]	[0.14]	[0.08]	
Poor nutrition	-0.13	0.21	0.16	0.10*	248
	[0.20]	[0.16]	[0.10]	[0.06]	
Happiness	-0.34	-0.59	-0.11	-0.33**	249
	[0.53]	[0.45]	[0.20]	[0.13]	
Days overdrawn	0.1	0.17	0.04	0.09**	241
	[0.13]	[0.11]	[0.06]	[0.04]	
Ln(net income)	-0.03	-0.12	-0.16**	-0.15***	240
	[0.18]	[0.12]	[0.08]	[0.06]	
Buffer saver	-0.23	-0.36	0.19	-0.12	248
	[0.48]	[0.41]	[0.24]	[0.17]	
Index: Bad	0.28	0.69***	0.32**	0.25***	226
	[0.27]	[0.22]	[0.14]	[0.08]	

Notes: OLS estimates. Each row reports coefficients from a regression with a different life outcome as the dependent variable. Key independent variables include indicators for choice patterns consistent with declining, increasing, and intransitive behavior in the experiment, as well as the average IRR across time horizons for an individual. Controls include gender, age, age squared, cognitive ability (standardized average of performance on two short-form tests of cognitive ability, see Dohmen et al., 2010, for more information), and an indicator for being highly educated (indicator for having taken the *Abitur*, a college entrance exam in Germany). BMI denotes body mass index; Smoker is an indicator for smoking more than occasionally; Drinking is an index for frequency of drinking (combining information about wine, spirits, beer, and mixed drinks), with higher values indicating more frequent drinking; Poor Nutrition is self-reported degree of poor nutrition, on a 4-point scale, with higher values indicating worse nutrition; Happiness measures subjective wellbeing on a 10-point scale with higher values indicating greater life satisfaction; Days overdrawn is days per year that current deposit account is overdrawn; Buffer Saver is an indicator for self-reported tendency to keep a savings reserve in case of unexpected expenses; Index: Bad is an index combining standardized values of all other outcome variables in the table. The sample excludes individuals who have missing values for the dependent variables, and individuals for whom discounting type, or level of IRR, are not unambiguous due to right-censoring in one or more time horizons. Results are similar using weaker conditions for censoring, i.e., excluding only if censored in two or more horizons. Robust standard errors in brackets; ***, **,* indicate significance at 1-, 5-, and 10-percent level, respectively.

A.2 Experiment Instructions

In the following we present a translation of the German instructions. Instructions were presented to the interviewer on the screen of the laptop computer, and were read aloud to the subjects by the interviewer.

Screen 1

Now that the interview is over we invite you to participate in a behavioral experiment, which is important for economic science. The experiment involves financial decisions, which you can make in any way you want to. The questions are similar to those asked in the questionnaire with the exception that THIS TIME YOU CAN EARN REAL MONEY!

I will first explain the decision problem to you. Then you will make your decisions. A chance move will then determine whether you actually earn money.

Every 7th participant wins!

HOW MUCH MONEY YOU WILL EARN AND AT WHICH POINT IN TIME WILL DEPEND ON YOUR DECISIONS IN THE EXPERIMENT.

If you are among the winners, your amount will be paid by check. In this case the check will be sent to you by post.

Screen 2

Participants were then shown a choice table for the respective experiment as an example. The table was printed on a green piece of paper and was handed to participants for them to study.

The experimenter continued explaining how the experiment would work.

The interviewer gave the following explanation:

In each row you see two alternatives. You can choose between

- A fixed amount of 100 Euro (column A "today")
- and a somewhat higher amount, which will be paid to you only "in 12 months" (column B).

Payment "today" means that the check you get by post can be cashed immediately.
Payment "in 12 months" means that the check you get can be cashed only in 12 months.

You start with row 1 and then you go down from row to row. In each row you decide between 100 Euro today (column A) and a higher amount (column B); please always keep the timing of the payments in mind. The amount on the left side always remains the same, only the amount on the right side increases from row to row.

Which row on one of the tables will be relevant for your earnings will be determined by a random device later.

Screen 3

As you can see, you can earn a considerable amount of money. Therefore, please carefully consider your decisions.

Can we start now?

If the participant agreed, the experiment started. If not, the experimenter said the following:

The experiment is the part of the interview where you can earn money! Are you sure that you DO NOT WANT TO PARTICIPATE?

If the participant still did not want to participate, the experiment was not conducted and the participant answered a few final questions. In case the subject wanted to participate the experiment began.

Participants studied their table. The experimenter asked for the subject's decision in each row, whether they preferred the option in Column A or B, starting with the first row. In case a participant preferred the higher, delayed amount the experimenter asked:

You have decided in favor of the higher amount of X in X months. We assume that this implies that for all higher amounts you also prefer the later payment, meaning that for all remaining rows all higher amounts will be selected (i.e., Column B).

If the participant did not agree, he kept on deciding between columns A and B.

Once the first table was completed, the second table was presented to the participant. The experimenter then said:

Now there is a second table. Please look at the table. You will do the same as before but please note that the dates of payment and also the payments on the right side of the table have changed.

For the second and third tables, the same procedure as with the first table was followed.

When the tables were completed, participants were asked whether they had thought about interest rates during the experiment and if so, which interest rate they had in mind and whether they had compared this interest rate with those implied in the decision tables. They were then asked what they would do with the 100 Euro from the experiment within the next weeks. Alternative answers were, "spend everything", "spend most of the money and save something", "save most of it and spend something" and "save everything", or "no reply".

Then it was determined whether the participant was among those who would be paid. Participants could choose their "lucky number" between 1 and 7. They could then press on one out of seven fields on the computer, which represented numbers from 1 to 7. If they hit "their" number they won, otherwise they did not win. In case they won, it was *determined which of the tables was selected and which row of the respective table. This was done again by pressing on fields presented to participants on the computer screen. In the end subjects who had won were informed that they would be sent the check by mail.*