Interpreting Time Horizon Effects in Inter-Temporal Choice

Thomas Dohmen[1], Armin Falk[2], David Huffman[3], Uwe Sunde[4]

December 2, 2011

Abstract

We compare different types of designs that have been used to test for an impact of time horizon on discounting. With the most commonly used type of design we replicate the typical finding of declining (hyperbolic) discounting, but with other designs reliably find constant or increasing discounting. Thus, the data as a whole are not well explained by any standard discounting assumption, and focusing on only one type of measure can be misleading. The choice pattern also implies a violation of transitivity, contrary to a broader class of models. Our results are based on real incentives, and replicate across two different representative data sets. Additional tests indicate that subjective perceptions and reference points play a role in time horizon effects. We discuss directions for future modeling of inter-temporal choice, implications for interpretation of previous evidence, and potential applications of reference-dependence to policy and understanding firm behavior.

Keywords: Time preference, hyperbolic discounting,
dynamic inconsistency, transitivity
JEL codes:

We would like to thank Matthew Rabin, David Laibson, Doug Bernheim, John List, Daniel Read, George Loewenstein, Colin Camerer, Steffano DellaVigna, Yoram Halevy, Stephan Meier, Charlie Sprenger, Lise Vesterlund, Johannes Binswanger, Roberto Weber, Tim Salmon, Thomas Vischer, and participants at various workshops and seminars for helpful comments. Excellent research assistance provided by Thomas Vischer, Jeff McManus, and Peter Akkies. Research was conducted in accordance with IRB guidelines.

[1] Research Centre for Education and the Labour Market (ROA), Maastricht University, IZA, and DIW; e-mail: t.dohmen@maastrichtuniversity.nl
[2] University of Bonn, IZA, and CEPR; e-mail: armin.falk@uni-bonn.de;
[3] Swarthmore College and IZA; e-mail: dhuffma1@swarthmore.edu
[4] University of St. Gallen, CEPR, and IZA; e-mail: uwe.sunde@unisg.ch.

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, showing that varying the time horizon affects choices in a way that suggests 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 dynamically inconsistent preference and self-control problems. Similar choice experiments have also been used by economists as a way to potentially measure the extent of dynamic-inconsistency among different groups of people, or predict economic outcomes thought to be related to self-control (e.g., Sutter et al., 2010, and many others).

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; Andersen et al., 2011; Sutter et al., 2010; Halevy, 2011). Also, studies relating experimental measures of dynamic inconsistency to 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 discounting assumption

[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, four studies use a different design, involving comparing non-overlapping time horizons of equal lengths (what we call a "shifted design" below). These find evidence that can be interpreted as declining discounting, but various design features of these latter studies might offer alternative interpretations, and make comparison to most 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; Harrison et al., 2010), or mixed evidence depending on the outcome

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 design and subject pool differences across studies.

This paper provides new evidence on how time horizon affects measured impatience, comparing different types of measures within the same framework. We adhere to the workhorse, "price-list" framework that is the industry standard for economists, but our approach is more comprehensive than most previous studies, in that we consider multiple time horizon lengths, starting dates, orders, and stake sizes, with nine treatments in all. To address concerns about subject pool and generalizability, we use two relatively large, representative samples of the adult population (N=500 and N=1,500), and investigate choice patterns for the overall population as well as robustness across various sub-populations of particular interest. Our experiments involve real money, with non-trivial stake sizes, and credibility of payments is very strong, for example because participants are in a long-term relationship with the surveying company and also the individual surveyor. We can check whether qualitative results are replicated across the two data sets despite different subjects, parameters, and timing of data collection.

We designed the experiment to explicitly compare three different approaches to testing discounting assumptions. 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 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.

being studied (Meier and Sprenger, 2008; Meier and Sprenger, 2010; Burks et al., 2011) or gender of the subject (Ashraf et al., 2006). 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.

[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.

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.

It is important to note these approaches to testing discounting assumptions typically involve a set of maintained assumptions, for example the assumption that people treat monetary payments as consumption. Relaxing these assumptions can affect the predictions of different models in important ways. We assess the ability of different models to explain the data, both with and without such assumptions.

The main stylized fact from our analysis is that time horizon matters for inter-temporal choice, in a way that is hard to explain with standard discounting models: People 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.[5] 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.

[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. The time horizon is a single day in both choices, but the start date is shifted into the future for the second choice. Declining discounting could explain a "preference reversal", such that the person prefers to have 1 apple in the choice involving today, but 2 apples in the choice about the more distant future.

[5] We discuss how, under some alternative assumptions, non-constant discounting models tend to make the same predictions as constant discounting. In this case, the models have trouble explaining why time horizon length matters. The same argument applies to quasi-hyperbolic models, if the "present-bias" falls in the brief time window before early payments arrive in our experiments.

[6] Transitivity is violated maintaining canonical assumptions of stationarity and separability. See Roelofsma and Read (2000) for some of the first evidence that inter-temporal choices can violate transitivity.

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). Conducting the analysis at the individual-level also shows that few people are consistently in line with any discounting model, while the majority exhibit the pattern observed at the aggregate level, which violates transitivity. The results thus pose an important puzzle for understanding inter-temporal choice.

The findings also provide a potentially useful lens through which to view previous evidence in the literature. They show that it can be misleading to focus on only one type of measure, and they raise questions about interpreting the results of studies that have this feature. They also show that "mixed evidence" can be generated by varying a particular aspect of the design, holding subject pool and other design characteristics constant. Indeed, the particular method variance we observe maps, broadly speaking, onto the patterns of inconsistent results observed across previous studies: we reliably find declining discounting with OD measures, similar to most previous studies using this approach, and less evidence for declining discounting using other types of measures, similar to many (but not all) previous studies using SD or OSD approaches.[7]

One alternative explanation for time horizon effects is that these partly reflect biased perceptions, rather than preferences (this type of argument has been proposed in various forms by Read, 2001, Rubinstein, 2003, Ebert and Pelec, 2007, and Zauberman et al., 2009).We do not observe perceptions directly, but one testable prediction of this explanation is that salient reference points might influence inter-temporal choice, given that subjective perceptions are well-known to be sensitive to such framing effects (e.g., Tversky and Kahneman, 1981).[8] Using one of our data sets, where treatment order was randomized, we find that the effect of time horizon length is still present, but smaller and not statistically significant, when comparing first treatments across subjects, while it

[7] Although our design is standard in many ways, and matches qualitative results in the previous literature, there are of course some differences relative to previous studies. For example, one strand of the literature has used early payments available literally on the day of the experiment, rather than a few days into the future as we (and other studies) do. While our results cannot speak to exactly how different time horizon comparisons would work in such a design, particularly in quantitative terms, in our view they at very least make a strong case for the importance of having more comprehensive designs with multiple time horizon comparisons, in order to check whether behavior is in fact explainable by some version of discounting models.

[8] In lab experiments with hypothetical payments, Zauberman et al. (2009) measure subjective perceptions of time duration directly, and find compression of perceived time duration. Ebert and Prelec (2007) show that framing effects alter perceptions of time duration in inter-temporal choice experiments (in the lab with hypothetical rewards).

is more pronounced and significant in later treatments. Thus, relative comparisons to a salient reference point, in the form of a previous treatment, do affect choice. Furthermore, time horizon effects are more pronounced when subjects are faced with such relative comparisons. Almost all previous studies use within-subject comparisons, and thus our results indicate that the effect of time horizon may be more about relative comparisons than has previously been realized.[9]

Our results are complementary to a strand of the literature in psychology, and also some recent work by economists, that questions whether time horizon effects reflect discounting. Previous studies (e.g., Barron, 2001; Read, 2001; Read and Roelofsma, 2003; Read et al., 2005) found, as in our study, "subadditivity" of discounting on average, such that discounting is more extreme when measured using sub-intervals.[10] Our study is complementary in that we use representative samples combined with incentivized experiments, assess the robustness of the results for different sub-populations, compare across a wide variety of different parameter values and time horizons lengths, show how relative comparison and reference points might be important for driving time horizon effects, replicate results across data sets, . Other studies have found evidence pointing to a role for subjective perceptions in explaining time horizon effects (Ebert and Prelec, 2007; Zauberman et al., 2009), but differ from ours in that they use hypothetical rewards and student subjects. Harrison et al. (2002) was the first to study discounting experiments in a representative sample, and Andersen et al. (2010; 2011) also do so; different from our study, both focus on OD comparisons.[11] Bisin et al. (2010) conclude in favor of a fixed-cost specification for quasi-hyperbolic discounting, focusing OD comparisons in the lab. 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

[9] Two studies that we are aware of allow comparing within-subject and between-subject designs, either implicitly or explicitly, and find results that are in line with ours, although they use hypothetical incentives. Ebert and Prelec (2007) find increased time sensitivity in a within-subjects design, and the data reported in Zauberman et al. (2009) indicate that the within-subject decline in discounting is about twice as large as the between-subject decline in discounting, although the authors do not comment on the difference.

[10] 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. Additivity is a prediction of all standard discounting models, regardless of the shape of the discount function.

[11] Assuming a specific functional form for utility, they use experimental measures to calibrate concavity of utility, and conclude that the resulting implied discount function is constant.

interview actually translate into the predicted direction of choice reversal when decisions are offered again in the future (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.

In summary, the findings have three main implications. First, the results pose an important puzzle for understanding inter-temporal choice, as the usual candidate models of time preference, and a broader class of models that requires transitivity, do not explain the modal choice pattern; existing models will of course continue to be useful in many settings, but given that the experiments involve real stakes and relatively simple tradeoffs, the inability to explain behavior is a notable limitation and calls for further research. At the end of the paper, we discuss some potential directions for modeling the behavior we observe. Second, our results suggest that designs focusing on only one type of measure might be misleading, by seeming to support one discounting assumption or another, and relatedly, variation in time horizon effects across populations, or correlations of such measures with economic outcomes, might not in fact reflect dynamic inconsistency or self-control problems, but rather something else. Third, the fact that the menu of savings options offered to an individual may influence inter-temporal choice has potentially important policy implications, even suggesting possible interventions designed to encourage saving; this reference-dependence may also help shed light on why firms present information about financial assets, or payment plans, to customers in the ways that they do.

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 offers evidence on alternative explanations. Section 5 concludes.

2 Design of the Experiment

2.1 Data collection and experimental procedures

Our analysis uses two data sets. One data set, denoted the SOEP data, involves a subsample of participants in the German Socio-Economic Panel (SOEP), a large panel data set for Germany (for a detailed description of the SOEP see Haisken-DeNew and Frick, 2003; Schupp and Wagner, 2002; Wagner et al., 2007). The second data set, denoted the Pretest data, 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 Pretest 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.[12] Both the Pretest and SOEP samples were constructed so as to be representative of the adult population, age 17 and older, living in Germany.[13] In total the Pretest data include 532 subjects, and the SOEP data include 1,503 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 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

[12] 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).

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

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 Pretest, 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 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 Tts is defined by starting date t, and ending date s. For a given horizon an individual chose between an early payment, X_t, available at the start of the horizon, or a larger, later payment, Z_s, available at the end of the horizon. In all choices for a given horizon, the amount of X_t was held constant, but the later payment, Z_s, was larger in each subsequent choice. For most time horizons, the value of Z_s 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_s implied an additional 2.5 percentage point increase in the annual rate of return, up to a maximum of 50 percent.[14] Some time horizons had 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_s, or equivalently the annual IRR, needed to induce the individual to wait. Across treatments, we varied t and s, and the amounts X_t and Z_s.

During the experiment subjects were presented with the different choices one at a

[14] 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.

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, and all subjects agreed.[15] 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, and thus would arrive within at most two days due to the well-known two-day guarantee for delivery by the German postal service.[16] Certificates for immediate payments could be cashed immediately, once they arrived, while certificates for payments in the future would be cashable only at the specified time.

In our design, there are several factors that helped to ensure equal credibility of early and late payments. This 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.[17] First, the design involved a "front-end delay" (see Coller and Williams,1999) in the sense that all payments arrived by mail between 1 and 2 days after the experiment, regardless of when the payment would be cashable. 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,

[15] As a consequence, there are no non-monotonic choices in the data. This procedure helped to mitigate potential confusion among subjects about the structure of the choice problem, by re-iterating the tendency for the late payment to only increase across further choices. One concern could be that this procedure tends to "lock-in" mistakes, if subjects make a mistake and then do not want to tell the experimenter that they have changed their mind. However, even if this were the case, to the extent that it affects all measures it should not matter for our conclusions, because they are based on differences across measures. Furthermore, studies that use alternative approaches and end up with non-monotonic choices are put in the position of needing to drop (non-randomly) a part of the data.

[16] 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.

[17] 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. For a series of reasons, discussed above, credibility is especially good 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" time horizon effects, this does not seem to explain the time horizon effects we observe.

about subjects expressing concerns regarding credibility of payments. Third, all participants in the SOEP data were members of the SOEP panel itself. Thus, these individuals were in a long-term relationship with the surveying agency, and typically even with the same individual surveyor, so that there should have been no doubts about credibility. Despite the fact that our setting created very strong credibility on the side of the agency conducting the experiment, a remaining potential concern might be that subjects worried that they themselves would misplace their payment before it became cashable, say, one year in the future. Empirically, however, we find no evidence that individuals had such concerns.[18]

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.[19] Making a trade-off between credibility and immediacy, we chose the shortest possible front-end delay compatible with avoiding a same-day credibility problem.[20]

[18] For example, the preference for early versus later payments is the same, regardless of whether the early payment is cashable immediately, or only in one year.

[19] 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.

[20] Mainly for the purposes of concreteness, so as to avoid having to say 1 or 2 days all the time, but also to heighten immediacy (without affecting credibility), 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",

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 Pretest data involved 500 subjects participating in the experiments, with three different measures of annual 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).[21] The order of the treatments in the Pretest data was randomized across individuals. The early payment was always 100 Euros, and the largest delayed payment always implied an IRR of 50 percent 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 52.5percent. 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 Pretest, even accounting for the lower payment probability of 1/9 rather than

or involves early payments in 12 months.

[21] As discussed above, 0 is a slight abuse of notation, since immediate payments arrived up to two days in the future.

1/7.[22]

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 qualitative predictions change if the assumptions are relaxed. Generally, relaxing these assumptions makes distinguishing constant and non-constant discounting not possible. This does not matter so much for our purposes, however, because we simply seek to assess whether there is some discounting model, or class of models, that can explain the data.

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 Pretest data. Assume compounding occurs once every 6 months, and for simplicity that a period is 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:

$$(1+\frac{r}{2})X_0 \lesseqgtr Z_6 \qquad (1+\frac{r}{2})^2 X_0 \lesseqgtr Z_{12} \qquad (1+\frac{r}{2})X_6 \lesseqgtr Z_{612}$$

The lowest delayed payment such that the individual prefers to wait establishes the points

[22] 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.

of indifference for each horizon, and defines the internal rates of return:[23]

$$(1+\frac{IRR_{T06}}{2})X_0 = Z_6 \qquad (1+\frac{IRR_{T012}}{2})^2 X_0 = Z_{12} \qquad (1+\frac{IRR_{T612}}{2})X_6 = Z_{612} \quad (1)$$

Solving for IRRs yields:

$$IRR_{T06} = 2(\frac{Z_6}{X_0} - 1) \qquad IRR_{T012} = 2((\frac{Z_{12}}{X_0})^{\frac{1}{2}} - 1) \qquad IRR_{T612} = 2(\frac{Z_{12}}{X_6} - 1) \quad (2)$$

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

$$(1+\frac{\gamma}{2})X_0 = Z_6 \qquad (1+\frac{\gamma}{2})^2 X_0 = Z_{12} \qquad (1+\frac{\gamma}{2})X_6 = Z_{12} \quad (3)$$

Where γ is the constant rate of time preference. As this is the same as condition (1), if IRRs are the same in all three horizons, it follows directly that a constant discounter chooses Z_s such that the measured IRR is invariant with respect to time horizon: $IRR_{T06} = IRR_{T012} = IRR_{T612}$.

2.3.2 Declining and increasing discounting

In the case of declining discounting, there are different discount rates, γ_1 and γ_2, for periods 1 and 2, such that $\gamma_1 > \gamma_2$. In this case the points of indifference in T06, T012, and T612 are given by

$$(1+\gamma_1)X_0 = Z_6 \qquad (1+\gamma_1)(1+\gamma_2)X_0 = Z_{12} \qquad (1+\gamma_2)X_6 = Z_{12}. \quad (4)$$

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

$$IRR_{T06} = 2((1+\gamma_1)-1) \quad IRR_{T012} = 2(((1+\gamma_1)(1+r_2))^{\frac{1}{2}}-1) \quad IRR_{T612} = 2((1+\gamma_2)-1) \quad (5)$$

[23] Because the delayed payment is a discrete variable in the experiment, the lowest delayed payment that is preferred actually establishes an upper bound for the IRR, while the largest delayed payment that is not preferred establishes the lower bound. Without any consequences for the qualitative results, we neglect this issue and use the upper bound when deriving predictions.

Given $\gamma_1 > \gamma_2$ this implies $IRR_{T06} > IRR_{T012} > IRR_{T612}$. Intuitively, impatience should be greatest in T06 with declining discounting, because it is the most focused on periods close to the present. Behavior in T612 should be the most patient, because it only concerns future payments. 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.[24] 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, *ceterus paribus.* Aside from predicting a magnitude effect, however, this model has the same properties as the quasi-hyperbolic

[24] 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: 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 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.[25]

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 going market interest rate. Outside of the experiment the 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.

If the assumption of linear utility is not accurate, and the unobserved utility function

[25] In this model, a payment Z_t received in the future is discounted by $\Delta_t = \delta^t - \frac{b}{Z_t}$, where $b > 0$ 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 $\Delta Z_t = \delta^t Z_t - b$. Importantly, when comparing two future payments, the fixed cost applies to both, and thus cancels out, so that between-period discounting in the future is the same as in the exponential model (just as in the quasi-hyperbolic model with variable costs). Maintaining the assumption that the present extends more than two days into the future, the indifference conditions implied by choices are given by

$$(1+\frac{\gamma}{2})X_0 + (1+\frac{\gamma}{2})b = Z_6 \qquad (1+\frac{\gamma}{2})^2 X_0 + (1+\frac{\gamma}{2})^2 b = Z_{12} \qquad (1+\frac{\gamma}{2})X_6 = Z_{12}$$

Substituting into (2) yields

$$IRR_{06} = 2((1+\frac{b}{X_0})(1+\frac{\gamma}{2}) - 1) \qquad IRR_{012} = 2((1+\frac{b}{X_0})^{\frac{1}{2}}(1+\frac{\gamma}{2}) - 1) \qquad IRR_{612} = 2((1+\frac{\gamma}{2}) - 1)$$

These equations imply $IRR_{T06} > IRR_{T012} > IRR_{T612}$, the same as hyperbolic, or regular quasi-hyperbolic, discounting models under these assumptions. 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.

is instead concave, this also affects predictions, in a way that depends on the particular time horizons being compared. Given that later payments are always larger, diminishing marginal utility of money would be another reason, besides time preference, why later payments are discounted relative to early payments. Unobserved concavity therefore leads to overestimating the degree of impatience, but importantly, the bias should be larger for long time horizons, because in that case payments are larger in absolute magnitude. Thus, for OD comparisons (where we find the strongest declining discounting), it works against finding declining discounting.[26] Analogously, for OSD comparisons (where we find the strongest increasing discounting) it tends to understate increasing discounting. SD comparisons should be unaffected, since the decline in marginal utility affects both measures in exactly the same way and is differenced out. Thus, while we do not identify concavity of the utility function in our design, the direction of the bias is clear, and it turns out that correcting for concavity would only strengthen our conclusions about aggregate patterns.[27]

If the assumptions about time separability and time stationarity of the utility function are not valid, then measured IRRs may vary 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, 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 existing models, maintaining assumptions of separability and stationarity, which are canonical and very widely used.

[26] If the estimates of impatience are affected by concavity of the unobserved utility function, we have $I\tilde{R}R_{T06} = 2(\frac{u(Z_6)}{u(X_0)} - 1)$, $I\tilde{R}R_{T012} = 2((\frac{u(Z_{12})}{u(X_0)})^{\frac{1}{2}} - 1)$, and $I\tilde{R}R_{T06} = 2(\frac{u(Z_{12})}{u(X_6)} - 1)$. By concavity, it is clear that $I\tilde{R}R_{T06} < IRR_{T06}$, $I\tilde{R}R_{T012} < IRR_{T012}$, and $I\tilde{R}R_{T612} < IRR_{T612}$, so that we overestimate the level of impatience in each case. It is also clear, however, that overestimation is more severe for T012 than the shorter horizons. Thus, we have $I\tilde{R}R_{T06} - I\tilde{R}R_{T012} < IRR_{T06} - IRR_{T012}$, and $I\tilde{R}R_{T612} - I\tilde{R}R_{T012} < IRR_{T612} - IRR_{T012}$, and we understate the difference in IRRs between short and long time horizons if utility is concave.

[27] When the goal is to estimate the precise level of the discount rate for a given horizon, rather than just an upper bound, one approach is to assume a specific utility function, and calibrate the curvature using risk aversion experiments (see Andersen et al., 2008; Andersen et al., 2010). For other approaches see, e.g., Attema et al. (2010), and Andrioni and Sprenger (2009).

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 Pretest 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.[28]

In the top panel of Figure 1, we see that people in the Pretest data are more patient for T012 than for T06, or T612. People are quite similar in impatience, however, for the T06 and T612 measures.[29] Thus, people are more impatient for shorter time horizons than longer time horizons, but similarly impatient over the short horizons regardless of 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, and even greater impatience for T01 than T06. 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.[30] 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.[31] Despite this framing difference between the one-month measures, they are all clearly more similar to each other, than to measures with longer horizons.

Table 2 reports the results in terms of descriptive statistics. We focus our discussion on median IRRs, as the right-censoring of the data makes interpretation of means more difficult, and likely understates the true impact of time horizon length.[32] The table shows

[28] The cumulatives for the two measures considered separately are very similar, and are not significantly different. Means (20.48 and 20.29) and medians (17.50 and 17.50) are also very similar.

[29] Distribution tests confirm that the 6-month measures were significantly different from the 12-month measure ($p < 0.001$;$p < 0.001$; Kolmogorov-Smirnov), but that the 6-month measures were not significantly different from one another ($p < 0.90$; Kolmogorov-Smirnov).

[30] The distributions for T06 and T012 are significantly different, as are distributions for T01 and T06 ($p < 0.001$;$p < 0.001$; Kolmogorov-Smirnov). The distributions for T01b and T1213 are not significantly different ($p < 0.98$; Kolmogorov-Smirnov).

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

[32] Means are calculated coding right-censored IRRs with a value of 52.25, or 105, depending on the upper bound for the time horizon measure. This top-coding likely understates the level of the true mean IRR for a given horizon. Importantly, however, as seen in the table, the fraction censored decreases

that median IRRs are higher for short than for long time horizons, without exception, in both Pretest and SOEP data sets. IRRs for time horizons of the same length, by contrast, are similar despite different starting dates. Differences are generally statistically significant, or not significant, in line with the results based on cumulative distributions.[33] Also as expected, given the differing upper bounds, medians for T01 in the second sub-sample of the SOEP data are substantially smaller than those for the T01b and T1213 measures in the third sub-sample.[34]

The impact of time horizon length on median IRRs is also substantial in magnitude, although magnitudes vary across data sets, potentially due to variation in stake sizes. For example, in the Pretest, both six-month horizons have IRRs that are at least 10 percentage points higher than the IRRs for the 12-month horizon (see Table 2). In the SOEP data, we see an even larger effect of time horizon length, comparing the one-month to twelve-month horizons. For example, the IRR for T01 is 25 percentage points greater than the IRR for T012. Interestingly, however, for T06 and T012 in the SOEP the effect of time horizon length is smaller than the comparable horizons in the Pretest: The difference between T06 and T012 is 2.5 percentage points rather than 10. The levels of IRRs are also lower for each time horizon.[35] Given that we increased stakes sizes in the SOEP, one interpretation is a "magnitude effect", such that the level of impatience, and impact of time horizon length, 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 (see Frederick et al., 2002). An alternative explanation, however, could be that the unobserved utility function is more linear in the range of stakes offered in the SOEP

monotonically with greater time horizon length, which suggests that the mean IRR is more strongly biased downwards for short time horizons. This implies that differences in mean IRRs tend to understate the true differences between mean IRRs for short and long time horizons. Indeed, the differences in mean IRRs across short versus time horizons are almost always smaller than differences in median IRRs, and this is especially true when comparing the longest to the shortest horizons, exactly as one would expect with greater downward bias in mean IRRs for shorter horizons.

33 Significance levels are based on paired t-tests, or (non-parametric) signed-rank tests, for equality of means and medians, respectively.

34 We may also understate the difference in median IRRs when comparing T01 to T012 in the SOEP data, because the median IRR in T01 is right-censored.

35 An exact comparison involves taking the sub-sample of individuals in the Pretest data who answered T012 first, and T06 second, because this is the order used in the SOEP. Median IRRs are 11.25 percentage points lower for T012 , and 15 percentage points lower for T06, comparing SOEP to Pretest. The difference between IRRs for T06 and T012 is 2.5 percentage points, rather than 6.25 as in the Pretest. The differences in the levels are statistically significant for both time horizons, pooling data from the Pretest and SOEP and using non-parametric rank-sum tests ($p < 0.01$; $p < 0.01$). The difference in the difference between the IRRs for T06 and T012 is not, however, statistically significant ($p < 0.46$).

(see Andersen et al., 2011, for a discussion).[36]

Tables 3 and Table 4 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 3 we see that IRRs in the Pretest 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 4 presents similar regression analysis based on the SOEP data. Results are reported separately for the three sub-samples. 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 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.

The regression analysis also allows checking robustness with respect to order effects, in the Pretest 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 3, and again find a significant difference between T012 and T06, but not T06 and T612. Furthermore, all interaction terms are not statistically significant, individually or jointly.[37] Thus, the qualitative results are systematic across all different possible treatment orders, and are not driven by any particular treatment ordering.

In subsequent columns, Table 3 and Table 4 explore the robustness of the qualitative 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

[36] One example of a model that can generate a magnitude effect is the fixed-cost version of the quasi-hyperbolic model (Benhabib et al., 2010).

[37] Results available upon request.

market interest rates, and potentially affect sensitivity to varying time horizon.[38] 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.[39] 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 degree of 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.[40] 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. To identify individual "types", we focus on the Pretest data, because we have seen that it is important to consider more than just one type of measure, and in that case there are multiple measures (OD, SD, and OSD) for each person. Table 5 reports fractions of individuals exhibiting different possible qualitative choice patterns across the different measures. Adhering to the usual maintained assumptions for testing discounting assumptions, we classify individuals who consistently have the same IRR across all time horizons as constant discounters (this fraction would include quasi-hyperbolic, or subjects engaging in arbitrage, under weaker assumptions). People with $IRR_{T06} > IRR_{T012} > IRR_{T612}$ are classified as declining discounters, and people with the opposite pattern are classified as increasing.

One feature of the individual-level data that is immediately apparent is that fractions of individuals who can be classified as being fully consistent with the predictions of constant, declining, or increasing discounters are relatively small. The first row of Table 5 reports these fractions, which are 13.37, 8.80, and 7.04 percent, respectively. The modal pattern, as reflected in the aggregate results, is for individuals to be more impatient in

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

[39] 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?"

[40] 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 Pretest 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.

both short time horizons than the long time horizon; almost 50 percent of individual exhibit this behavior. Among these, roughly half have equal IRRs for the two short time horizons, exhibiting zero sensitivity to starting date, while the others have different IRRs for the two short horizons. The remaining fraction of the sample, in the "Other" category, is also inconsistent with discounting assumptions. In 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 70 percent of individuals are inconsistent with all discounting types.

Classification of types is less clear if people make "trembling hand" errors. To assess how the classifications change if there are in fact such errors, we converted each observed IRR into an interval 5 percentage points wide, corresponding to an error of plus or minus 2.5 percentage points in either direction. We then classified IRRs are being different only if their respective intervals do not overlap at all. The second row of Table 5 shows the results. Not surprisingly, the fraction of constant discounters increases substantially, to 36.92 percent, using this conservative approach that favors finding constant discounting. The fractions of declining and increasing discounters are only 5.02 and 4.30 percent, respectively. The fraction of individuals who are inconsistent with standard discounting models remains substantial, at 53.77 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 some 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 Pretest 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.

The pattern is also inconsistent with non-constant discounting models where the discount rate changes discretely in the first one or two days (if the change occurs farther in the future, qualitative predictions are the same as declining discounting models discussed above). Such models would predict constant IRRs across time horizons, because even the earliest payments are available starting two days in the future. This class of models includes the quasi-hyperbolic model, 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).[41] The finding that time horizon does strongly influence IRRs is thus inconsistent with this class of non-constant discounting models.

Relaxing the assumption that people treat money as consumption, or allowing for concavity of utility, 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. Accounting for concavity of the unobserved utility function, the pattern we observe would become even more pronounced, leaving conclusions unchanged.

We summarize the ability of different models to explain individual time horizon comparisons, as well as the data as a whole, in Table 6. 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. This shows how focusing on a particular type of time horizon comparison may be misleading, by seeming to support one class of models or another.

Maintaining only the usual workhorse assumptions of separability and time stationarity of the period utility function, the data are inconsistent with a broader class of models because they imply a violation of transitivity. For example, T06, T012, and T612 can all be represented as choices between triples, where elements denote utilities at time 0, 6, and

[41] The recent two-system model of Fundenberg and Levine (2010) makes similar predictions to the hyperbolic model, allowing for a continuous decline in discount rates moving into the present. Thus, the model is not fully consistent with the data, for same reasons discussed in the context of declining discounting models.

12.

$$\begin{array}{llllll} & T06 & & T012 & & T612 \\ A & (\Delta_0 u(X),0,0) & A & (\Delta_0 u(X),0,0) & D & (0,\Delta_6 u(X),0) \\ B & (0,\Delta_6 u(Y),0) & C & (0,0,\Delta_{12} u(Z)) & E & (0,0,\Delta_{12} u(Y)) \end{array} \tag{6}$$

where Δ_t is a potentially time-varying discount factor, indexed by distance from the present (calendar date) and $u()$ is the (stationary) instantaneous utility function. In the data, for $X < Y < Z < 2Y$ (more precisely, for X=100, Y=116, Z=128) we observe $A \succ B$, $C \succ A$, and $D \succ E$. Because we observe $D \succ E$, we know that $\Delta_6 u(X) > \Delta_{12} u(Z)$. We now consider what this implies for hypothetical choice T612', corresponding to

$$\begin{array}{ll} & T612' \\ B & (0,\Delta_6 u(Y),0) \\ C & (0,0,\Delta_{12} u(Z)) \end{array} \tag{7}$$

Assuming that $u()$ is weakly concave, $\Delta_6 u(X) > \Delta_{12} u(Y)$ implies $\Delta_6 u(Y) > \Delta_{12} u(Z)$. This follows directly from (potentially weak) concavity and the fact that $X < Y < Z$, and $Y - X > Z - Y$. In other words, in T612' we will have $B \succ C$, which implies $B \succ C \succ A \succ B$ and a violation of transitivity. This violation is present for *any* arbitrary discounting function.[42]

In summary, the data pose a puzzle for modeling inter-temporal choice. The choice pattern is not well explained by the usual candidate models of inter-temporal choice, or by a broader class of models that may have any form of discount function, and any weakly concave utility function, but require transitivity.[43] Furthermore, the choice pattern is very robust: We see the same patterns in the population as a whole, as well as within different sub-populations, we replicate the results across different data sets, and we find the results despite using real incentives.

[42] The choice pattern labeled "subadditivity" by Read (2001) also implies a violation of transitivity, although in that study the connection between additivity and transitivity is not made explicit.

[43] Without the usual restriction of time stationarity, the period utility function can vary systematically with distance from the present, and the data need not imply a violation of transitivity. To see this, note that $\Delta_6 u_6(X) > \Delta_{12} u_{12}(Y)$ need not imply $\Delta_6 u_6(Y) > \Delta_{12} u_{12}(Z)$, for example if $u_6(Y) - u_6(X) < u_{12}(Z) - u_{12}(Y)$. Relaxing the usual assumption of a separable discount function would also allow rationalizing the data without a violation of transitivity, as would allowing the discount function to be indexed along some dimension other than distance in time from the present (calendar date).

4 Reference comparisons and subjective perceptions in inter-temporal choice

We have seen that inter-temporal choices in the experiments are hard to explain with discounting models, and in terms of the objective stakes and time durations, choices even imply violations of transitivity. In this section we investigate one potential alternative mechanism that could be important for explaining the behavior: A deviation of subjective perceptions about time duration and monetary magnitudes, from objective magnitudes. Some previous work, mainly in psychology, has pointed specifically to biased perceptions of time duration as a factor in time horizon effects (this type of explanation has been proposed in various forms by, e.g., Read et al., 2001; Rubenstein, 2003; Ebert and Prelec, 2007; Zauberman et al., 2009). This type of bias could explain some of the main patterns we have observed. For example, if doubling objective time duration leads to less than double the subjectively perceived duration, this creates a tendency, all else equal, to be more patient for long time horizons that short horizons, independent of starting date. Subjective perceptions of monetary magnitudes could also play a role, however, as different time horizons involve different absolute payoff amounts, despite keeping the annual rate of return constant.

We do not measure subjective perceptions directly, but this explanation generates testable predictions, in the sense that salient reference points are well known to influence subjective magnitude judgments (Tversky and Kahneman, 1981)). We exploit the randomization of treatment order in the Pretest data to investigate whether the absence or presence of a salient reference point, in the form of an earlier treatment, influences subsequent choice behavior. A salient comparison to another time horizon could influence both perceived time duration, as well as magnitude of monetary amounts. If sensitivity to time horizon length is more pronounced for later treatments, rather than the first, this suggests that subjective perceptions of *relative* magnitudes are an important underlying mechanism, and also shows that reference points can influence inter-temporal choice (discounting models do not predict any effect of reference points).

Figure 2 reproduces the cumulative distribution functions for the Pretest data, but using only first treatments in the left-hand panel, and second and third treatments in the middle and right-hand panels, respectively. We see that the stylized facts discussed in relation to Figure 1 are more pronounced, and estimated more precisely, in later treatments. Looking at first treatments, across subjects, there is a smaller impact of time horizon on IRRs: The median IRR is 2.5 percentage points lower for T012 than T06, and 7.5 percentage points lower than for T612. These differences are statistically significant ($p < 0.46$; $p < 0.19$).[44] By contrast, there is is a larger time horizon effect for both second and third treatments, with the IRR for T012 always at least 15 percentage points lower than the IRRs for either six-month horizon. Differences are statistically significant, for both second ($p < 0.01$; $p < 0.01$) and third treatments ($p < 0.01$; $p < 0.01$).[45] Notably, we found previously that the effect of time horizon length is still observed when controlling for different possible orders. Thus, while the tendency to be more patient for long horizons is more pronounced with relative comparisons, the pattern is systematic and emerges for all treatment orders.

In summary, while we still find higher median IRRs for short than long time horizons in first treatments, relative comparisons appear to play an important role in making these time horizon effects more pronounced, and more precisely estimated. These findings are consistent with a role for subjective perceptions in time horizon effects, which are influenced by relative comparisons. In particular, they suggest that valuations of monetary payments are reference dependent. The contrast between a long and short time horizon would seem to make the duration in the long horizon seem longer, which works against an increase in patience for long horizons; our results can be explained by the contrast along the monetary dimension being more important, as the contrast with the smaller delayed payments in short horizons make the delayed payments in long horizons seem relatively more attractive, enhancing willingness to wait.

These findings indicate that time horizon effects may be more about relative comparisons, along the money dimension, than has previously been realized. They also point to the potential importance for inter-temporal choice, of the menu of time horizon options presented to individuals, in terms of the presence or absence of explicit reference comparisons.

[44] Results are based on non-parametric rank-sum tests.

[45] Results are based on non-parametric rank-sum tests.

5 Conclusion

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 different sub-groups, and it replicates across different data sets and different types of measures.

For theory, this puzzle is a call for further research on modeling inter-temporal decision-making. 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. Yet, discounting models, regardless of whether discounting is constant or non-constant, miss some systematic and robust features of observed behavior. This points to an opportunity, to improve models of inter-temporal decision making, in a way that may prove useful not just in the context of experiments but also in the field.

One potential way to model our findings would be allow for discounting that is indexed not by distance from the present (calendar date) but rather, say, time delay between early and late payments, independent of calendar date. A simple version of such a model would involve a person requiring a fixed amount k to be willing to pass up an early payment for a later payment, regardless of the timing of payments relative to the present.[46] In this case, while a given IRR might not be sufficient to induce waiting for a short horizon, because the absolute monetary gain, Δ_i, is less than k, the same IRR could induce waiting for a longer horizon, because in that case the IRR necessarily involves a larger delayed payment and the absolute gain, Δ_j, might be sufficient to exceed the fixed cost, k.[47] A fixed cost of waiting is also consistent with some other aspects of our results. Measured IRRs should decrease with stake size, and the difference in IRRs across short and long horizons should be smaller with larger stakes, as we observe. Also the difference in

[46] See Ebert and Prelect (2007) for a related discussion, on duration-dependent discounting, and Scholten and Read (2006, 2010) for alternative, attribute-based models of inter-temporal choice, which can explain similar findings.

[47] The fixed-cost version of quasi-hyperbolic discounting discussed in Benhabib et al. (2010) is different, because in that case discounting is still indexed by calendar date.

IRRs between short and long time horizons should also increasing in the disparity in time horizon lengths, consistent with the fact that T01 vs. T012 involves a greater disparity in IRRs than T06 vs. T012.

It is also possible to include some reference-dependence in such a model, to account for the more pronounced impact of time horizon length when individuals face an explicit comparison. For example, the perceived gain from waiting in a given time horizon is either Δ_j if there is no salient comparison to another time horizons, or else $\frac{\Delta_j}{v(\Delta_i)}$, where $v()$ is some increasing function, and Δ_i is the absolute gain offered previously for another horizon. In this latter case, the perceived gain from waiting for a long time horizon increases, having seen a short time horizon first, and vice versa, which can lead to the effects of relative comparisons observed in our data.

Such a modeling approach raises interesting possibilities for understanding inter-temporal choice more generally. For example, one implication that could be drawn from the model is a difference in decision making for low versus high stakes. The model implies that time horizon effects become smaller, and the level of IRRs are reduced, potentially to zero, the larger are the stakes involved in inter-temporal decisions. Thus, time horizon effects, and impatience, are more likely for the smaller-stakes decisions in life, while for large-stakes decisions individuals tend to be patient and insensitive to time horizon length. One important caveat about this generalization, however, concerns the magnitude of time horizon durations. If life decisions that involve high stakes also involve time horizon lengths that are quite long, on the order of, say, 10 or 20 years it could be that the fixed cost of waiting increases, relative to what we measure with the 1 year horizons in our experiment. In this case, the types of time horizon effects we observe in the experiments could persist even with very large stakes decisions.[48] Further investigation of how people perceive both the costs, and gains, from waiting, appears to be a useful direction for future research.

For the empirical literature on testing discounting assumptions, our results raise questions about how to interpret evidence on time horizon effects in previous studies. This is particularly true in the case of studies that use only one type of time horizon

[48] Some laboratory evidence also suggests that, while increasing stakes eliminates the pattern of higher IRRs for short time horizons, it instead leads to strong patterns of "superadditivity", such that people have *lower* IRRs for short time horizons than long time horizons, independent of starting date (Scholten and Read, 2006). Thus, time horizon effects first go to zero, but then start increasing again, as stakes increase. While this evidence is based on hypothetical stakes, it at least suggests the possibility of a non-monotonic relationship between stake size and time horizon effects.

comparison, and conclude in favor of one type of discounting assumption or another. Our findings illustrate how having only one type of measure can be misleading, because a more comprehensive design might reveal that the behavior is actually not explainable by any type of discounting function at all. Also, our findings indicate that the almost universal approach of using within-subject designs may conceal an important reference-dependent aspect of time horizon effects. Of course, while our design is quite standard, it does not capture all design features of all previous studies, or subject pool characteristics, and thus cannot speak to exactly how different measures would work under such conditions, particularly in quantitative terms. At the very least, however, the findings of our paper make a strong case for the need to run more comprehensive designs, as a check, before interpreting results as supporting a particular discounting model.

Our findings are also potentially relevant for interpreting the literature on using experimental measures of self-control problems to predict life outcomes. The typical methodology involves classifying people as declining or increasing discounters, based on one or sometimes two types of time horizon comparison (OD, SD, or both). Our results raise the possibility, however, that a substantial portion of individuals categorized in any one of these discounting categories may be misclassified. For example, someone who appears to be a constant discounter based on one type of time horizon comparison can actually be intransitive. A more correct specification would then involve an indicator for individuals who are intransitive, as distinct from (the perhaps relatively few) individuals who consistently match a given discounting assumption. Otherwise, what is being attributed to a given type of discounting function might instead reflect some other mechanism entirely. This potential for mis-specification raises questions about how to interpret results based on the usual methodology.

In our own data, we have also done some analysis relating experimental measures to outcomes. We constructed various measures of deviations from constant discounting, as well as measures of the level of impatience (e.g., average IRR for an individual across time horizons) and related these to smoking, poor nutrition, drinking, body mass index, low wealth, overdrawn checking account, self-reported tendency to save, and self-reported problem with overspending (results are available upon request). These types of outcomes have been examined in previous studies, but typically a given study has considered only a few rather than the whole set. Also different from previous studies, where possible we used a specification with an indicator for intransitive types (this was possible for our Pretest data). Our findings are as follows: to the extent that they are related to outcomes, deviations from constant discounting are associated with worse outcomes; qualitative results are almost identical for declining and intransitive types, in terms of which outcomes are predicted, and in which direction, although intransitivity is related to more outcomes; increasing discounting predicts little, but when it does, in the same direction as declining discounting or intransitivity; for the majority of the outcomes, there is no significant relationship to any of the measures, even more so after adjusting standard errors for multiple hypothesis testing. On the one hand, these results are not supportive of the qualitative predictions of discounting models, where declining and increasing discounters would be expected to have opposite types of self-control problems, over-indulging or under-indulging, respectively. Previous studies have also never found evidence that increasing discounters have the opposite behavioral patterns of declining discounters, and have typically found mixed predictive success as we do. Although declining discounting does predict outcomes, the fact that intransitivity, and increasing, do so in the same way, suggests that the relationship to outcomes might be due to a common underlying mechanism, as opposed to the different patterns picking up distinct types of discounting functions. Ultimately, the relationship with self-control outcomes is rather weak, which is not unexpected in light of our earlier findings, which cast doubt on interpreting time horizon effects as being related to self-control problems.

Interestingly, we did find relatively better predictive success when using the average IRR for an individual across time horizons (or other measures that capture the level of discounting rather than cross-horizon differences). This type of measure was a reasonably successful predictor of outcomes in both data sets, after corrections for multiple hypotheses. For example, poor nutrition, smoking, having an overdrawn checking account, and low wealth, are all more likely for someone with a higher IRR.[49] Thus, the results on outcomes are encouraging in that experimental measures do appear to provide a behaviorally meaningful index of impatience.

Based on our results, focusing on different aspects of the experimental design, or

[49] These findings 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). In Vischer et al. (2011) we reach similar conclusions about the predictive power of the level of impatience, based on relating the T012 measure in the SOEP to various economic outcomes.

developing new types of designs altogether, may be useful directions for future research. One possibility is to explore more systematically, within monetary choice experiments, how inter-temporal tradeoffs are influenced by the salience of different alternatives. Another possibility is to study whether behavior is more compatible with 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). 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 puzzling patterns of intransitivity as we find. If the goal is to capture self-control problems, other approaches might be also useful, for example studying how willingness to pay varies with the degree of salience, e..g, in terms of being able to see the good (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 economic outcomes.[50]

[50] See also an interesting, alternative paradigm for monetary choice experiments, holding monetary magnitudes constant and varying delay length, proposed by Noor (2011).

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 (2008): “Eliciting risk and time preferences,” *Econometrica*, 76(3), 583–618.

——— (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 (2009): “Risk preferences are not time preferences,” UCSD working paper.

——— (2010): “Estimating time preferences from convex budgets,” *NBER Working Paper No. 16347*.

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.

Attema, A., H. Bleichrodt, K. Rohde, and P. Wakker (2010): “Time-tradeoff sequences for analyzing discounting and time inconsistency,” *Management Science*, 56(11), 2015.

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

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.

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,” *The American Economic Review*, 100(4), 1556–1571.

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.

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.

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," *The American Economic Review*, pp. 1449–1476.

——— (2010): "Timing and Self-Control," *Levine's Working Paper Archive*.

Haisken-DeNew, J. P., and J. Frick (2003): *Desktop Companion to the German Socio-Economic Panel Study (SOEP)*. DIW, Berlin.

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*, pp. 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.

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.

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

Meier, S., and C. Sprenger (2008): *Discounting financial literacy: time preferences and participation in financial education programs*. IZA.

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

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*, pp. 185–199.

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," .

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.

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.

Scholten, M., and D. Read (2006): "Discounting by intervals: A generalized model of intertemporal choice," *Management Science*, 52(9), 1424.

——— (2010): "The psychology of intertemporal tradeoffs.," *Psychological review*, 117(3), 925.

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.

Sutter, M., M. Kocher, D. Rutzler, and S. Trautmann (2010): "Impatience and uncertainty: Experimental decisions predict adolescents field behavior," *Discussion Papers in Economics*.

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

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

Tversky, A., and D. Kahneman (1981): "The framing of decisions and the psychology of choice," *Science*, 211(4481), 453.

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

Figure 1: Cumulative Distributions of IRR

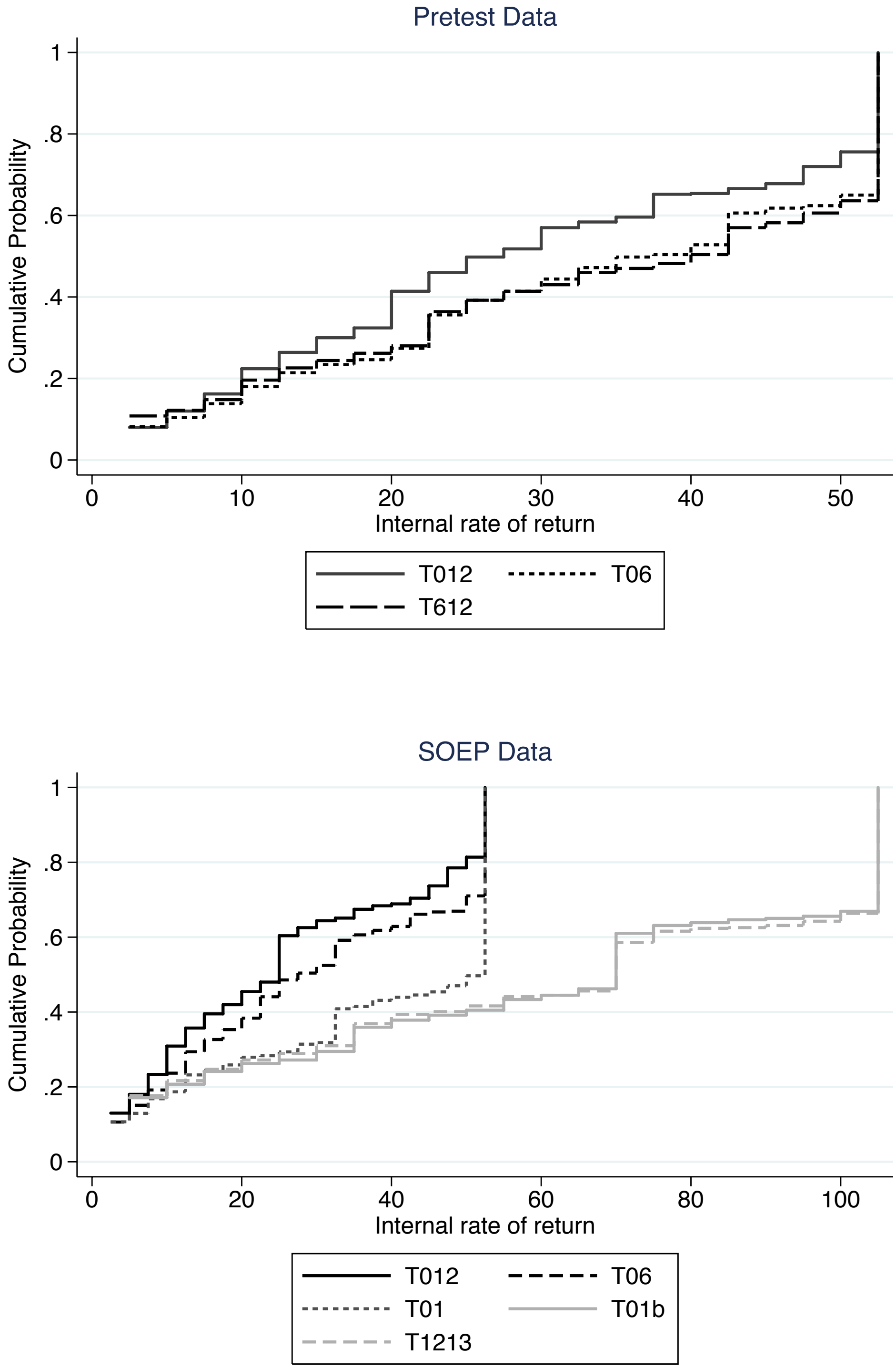
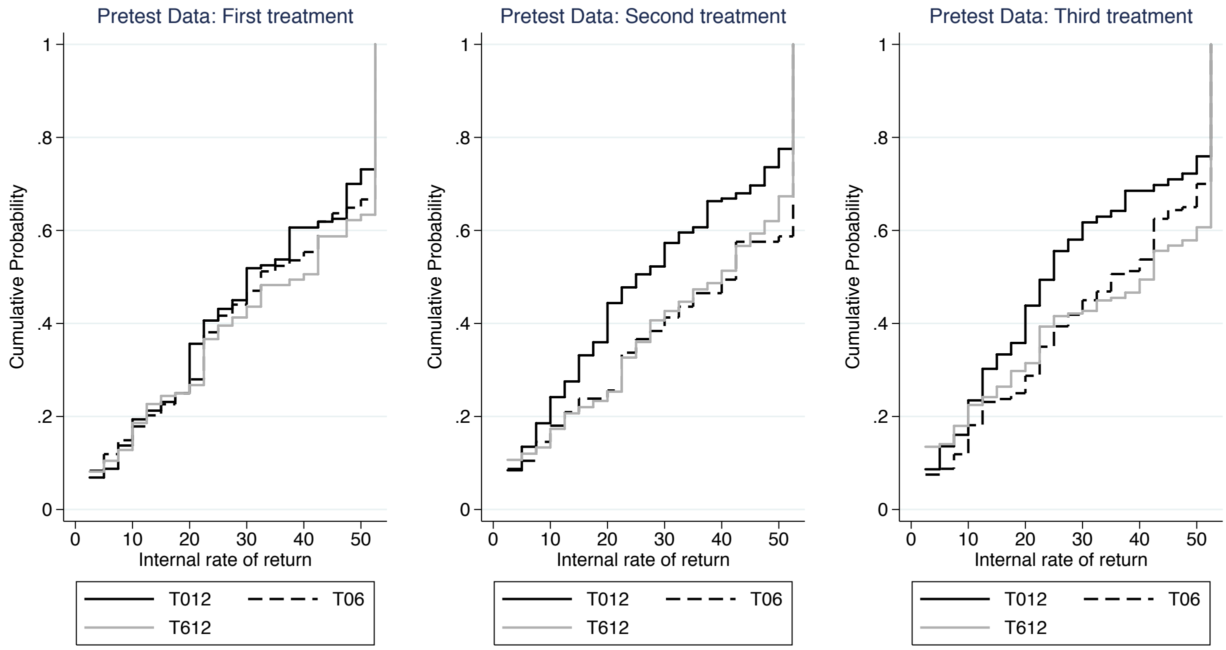


Figure 2: Treatment Order and IRR: Pretest Data



Tables

Table 1: Summary of Treatments

Measure	Data set	Sub-sample	Early payment (in Euro)	Upper-bound IRR	Obs.
T012	Pretest	n.a.	100	52.5%	500
T06	Pretest	n.a.	100	52.5%	500
T612	Pretest	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: Descriptive Statistics for IRR by Time Horizon

Measure	Data set	Mean IRR	Median IRR	S.D. of IRR	Fraction right-cens.		Mean test	Med. test
T012	Pretest	29.40	27.50	18.19	0.24	T012 vs. T06	$p < 0.01$	$p < 0.01$
T06	Pretest	33.56	37.50	18.10	0.35	T06 vs. T612	$p < 0.78$	$p < 0.05$
T612	Pretest	33.76	40.00	18.78	0.36	T612 vs. T012	$p < 0.01$	$p < 0.01$
T012	SOEP	26.07	25.00	18.41	0.19	T012 vs. T06	$p < 0.01$	$p < 0.01$
T06	SOEP	29.62	27.50	18.89	0.29	T06 vs. T01	$p < 0.01$	$p < 0.01$
T01	SOEP	36.56	52.50	19.41	0.53	T01 vs. T012	$p < 0.01$	$p < 0.01$
T01b	SOEP	60.87	70.00	39.08	0.33	T01 vs.T01b	$p < 0.01$	$p < 0.01$
T1213	SOEP	60.89	70.00	39.83	0.34	T01b vs. T1213	$p < 0.99$	$p < 0.88$

Table 3: IRR as a Function of Time Horizon, by Demographic Groups: Pretest 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)
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, with three time horizons (observations) per individual. Age groups are defined as less than or equal to the median age, and greater than median age, respectively. Low and high IQ indicate below and above median cognitive ability respectively. Less and more educated indicate whether an individual did not, or did, complete the Abitur, a college entrance exam in Germany. Low and high income indicate below and above median household income, respectively. 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: 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)
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
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
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. Low and high IQ indicate below and above median cognitive ability respectively. Less and more educated indicate whether an individual did not, or did, complete the Abitur, a college entrance exam in Germany. Low and high income indicate below and above median net personal income, respectively. 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 5: Individual types, Pretest 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	31.21	8.48	7.88	18.48	9.70	24.24

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 $IRR + / - 2.5$, and we count two horizons as being equal if their intervals overlap. The same restriction applies, in terms of excluding censored observations, but 14 additional individuals are excluded (N=330). Of the excluded individuals, 4 exhibit "cycling" or intransitivity in the relations between the three horizon measures, due to the ways in which intervals do or do not overlap. E.g., interval for T06 overlaps with interval for T012, but is strictly above interval for T612, while interval for T012 overlaps with interval for T612. This implies intransitivity because $IRR_{06} = IRR_{012}$, $IRR_{06} > IRR_{612}$, but $IRR_{012} = IRR_{612}$. The remaining 10 that are excluded involve the IRR for one time horizon being right censored, while the IRRs for both of the others are lower. Once we allow for errors, one or more of the lower horizons overlaps with the censored observation, and thus are no longer unambiguously classifiable.

Table 6: Stylized Facts on IRR and Time Horizon, and Different Models

	Constant discounting	Declining discounting	Increasing discounting	$\beta - \delta$, two-system, or "fixed cost" discounting, with present$\leq$ 2 days	Any discounting with arbitrage
Pretest 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. Discounting with arbitrage refers to predictions when individuals do not treat monetary payments as consumption.

A 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.*