The Robustness and Pervasiveness of Sub-Additivity in Intertemporal Choice

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

July 18, 2017

Abstract

A large literature has used choice experiments involving time-dated monetary rewards, to test whether time discounting is exponential or hyperbolic, with mixed results. One explanation, proposed by the psychologist Daniel Read (2001), is that the observed choice patterns reflect a type of framing effect, known as sub-additivity, rather than hyperbolic or exponential discounting. An alternative explanation, however, has emerged from a recent literature in economics, which points out various confounds that might affect the traditional intertemporal choice experiment, as well as challenges of inference from typically small or idiosyncratic samples. This paper makes two main contributions: (1) It re-visits the sub-additivity hypothesis, but using a design that addresses the key methodological confounds; (2) it uses large representative samples, to assess the pervasiveness and importance of anomalous choice patterns like sub-additivity. The analysis finds intertemporal choices that are consistent with sub-additivity, and rules out explanations based on confounds. Furthermore, sub-additivity is pervasive, being observed across all sub-populations studied in the analysis, and constituting the majority choice pattern at the individual level. The results underline that sub-additivity is an important feature of intertemporal choice, they raise caveats about how intertemporal choice experimens have often been interpreted, and they suggest some directions for methodological improvements.

Keywords: Time preference, hyperbolic discounting, self-control, dynamic inconsistency, intransitivity, sub-additivity.
JEL codes: D01, D90, D03, E21

[1] University of Bonn, ROA, IZA, and DIW; e-mail: tdohmen@uni-bonn.de;
[2] University of Bonn, briq, IZA, and CEPR; e-mail: Armin.Falk@briq-institute.org;
[3] University of Pittsburgh and IZA; e-mail: huffmand@pitt.edu;
[4] University of Munich, CEPR, and IZA; e-mail: uwe.sunde@econ.lmu.de.

1 Introduction

A large literature has used choice experiments to test whether time discounting is exponential or hyperbolic, but results are mixed. The typical experiment has involved offering individuals choices between earlier and later monetary rewards. Most of the early studies, conducted by psychologists, compared discounting over short and long time horizons, starting on the same date, for example discounting between today and 6 months, versus discounting between today and 12 months; in this paper this is denoted an Overlapping Horizons (OH) design. These early studies typically found that discount rates declined with time horizon length, consistent with, e.g., a hyperbolic discount function (for a survey see Frederick et al., 2002). The more recent literature, however, has found more varied results. Many of the recent studies have used a different design, comparing discounting over two non-overlapping horizons of the same length, for example, discounting from today to 6 months, versus 6 months to 12 months; in this paper this is denoted a Shifted Horizon (SH) design. Studies using SH designs have found various patterns, ranging from declining, to constant (exponential), to increasing (anti-hyperbolic) discounting.[1] Studies using experiments to classify individuals' discounting types have also found mixed evidence, in terms of whether indicators for declining discounting have the predicted relationship to behavior patterns such as procrastination, demand for commitment devices, or worse financial outcomes.[2]

One potential explanation for the mixed results has been proposed by the psychologist Daniel Read and co-authors.[3] Their experimental designs nested OH and SH comparisons, and examining this more holistic picture, they concluded that the impact of

[1] E.g., Kirby and Herrnstein, 1995 (and others) have found declining discounting, but Anderhub et al. (2001) and Sutter et al. (2013) (and others) have found constant or increasing discounting.

[2] Publication bias may have suppressed some null results, but various published studies find null or mixed results: Ashraf et al. (2006) find that an indicator for present bias predicts demand for commitment devices among women, but not among men; Sutter et al. 2010 find a null result when relating risky behavior among youths to indicators of present bias; various other studies find mixed evidence depending on the outcome being studied, e.g., a relationship of present bias to smoking but none for BMI or credit score (Meier and Sprenger, 2010; Burks et al., 2012; Meier and Sprenger, 2013); 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; with a more parametric approach, Harrison et al. (2010) find that smokers and non-smokers are "no more or no less dynamically inconsistent than each other" (pp. 716); Augenblick et al. (2015) find a null result regarding monetary indicators of present bias predicting demand for commitment devices, but Martinez et al. (2017) find that a monetary indicator for present bias predicts procrastination behavior.

[3] See, e.g., Read, 2001, Read and Roelofsma, 2003; Read et al., 2005; see also Baron, 2000.

varying time horizon on choices is not consistent with any standard discounting model. Specifically, asking for discounting over a 24 month interval led to less total discounting than asking about the same time horizon in terms of three 8 month sub-intervals. This suggests a type of framing effect, such that individuals are more impatient in the sub-interval frame, which Read labeled "sub-additivity."[4] Read argues that if the variation of choices with time horizon reflects sub-additivity, attempts to interpret behavior in such experiments in terms of hyperbolic or exponential discounting may be ill-founded.

An alternative explanation for the puzzle of mixed results, however, has emerged from a more recent literature in economics. This literature raises the possibility of various confounds that could cloud the ability to infer time preference from experiments using time-dated monetary rewards. One concern is that delayed monetary payments might be viewed as non-credible by subjects, therefore confounding attempts to infer discount rates from choices over rewards received at different times (Coller and Williams, 1999; Andreoni and Sprenger, 2012). Another concern is that concavity of utility could be confounded with time preference, since larger, later rewards would be discounted not just due to time preference but also due to diminishing marginal utility (Andersen et al., 2010 and 2011; Andreoni and Sprenger, 2012). A different issue is that time preference is about consumption rather than money, and thus individuals might engage in arbitrage between laboratory and market interest rates; in this case, choices over monetary rewards might reveal lending and borrowing opportunities rather than time preference (Cubitt and Read, 2007). The standard assumption that preferences are stationary over time has also been challenged, e.g., because credit constraints and unexpected shocks might make trade-offs change over time (Halevy, 2015). Finally, much of the previous literature has used either small, or non-representative, samples including research on sub-additivity; this might contribute to mixed results, and also leaves the importance of anomalous choice patterns like sub-additivity, in the general population, an open question.

The first main contribution of this paper is to revisit the sub-additivity hypothesis, but in a design with features that can help address the methodological concerns raised in the recent literature. Specifically, the setting is one in which experimenter and subjects are in a long-term relationship, leading to high credibility of (real) payments, and the design

[4] The notion of "sub-additive pricing" was discussed earlier by Kahneman and Knetsch (1994); they showed when a public good is framed in terms of constituent parts, willingness to pay for one these sub-items can be as great as for the good as whole.

also varies the payment procedure across subjects, allowing a robustness check regarding the issue of payment credibility. Calibration arguments, and a survey measure of risk aversion, allow checking whether concavity of utility might plausibly drive the observed patterns in inter-temporal choice. A post-experiment survey asks subjects whether they thought about market interest rates, a pre-requisite for arbitrage strategies, and allows investigating whether the results are driven by the minority who do think about interest rates. To assess a role for non-stationary preferences, the design allows checking whether results are similar for a variety of different time horizons involving different calendar dates, and includes an indicator for credit constraints, to directly check on that potential source of non-stationarity.

The second main contribution is to test for sub-additivity in representative samples. The experiments were conducted with two different representative samples of the German adult population ($N = 500$ and $N = 1,503$). Representative data allows assessing the prevalence and importance of potential sub-additive choice patterns: Whether these are confined to particular sub-populations, or whether this is a general feature of intertemporal choice.[5] The data also allow exploring how sub-additivity or other choice patterns are related to characteristics at the individual level, ranging from demographics, to cognitive ability, to financial situation.

The findings are consistent with sub-additivity being a fundamental feature of intertemporal choice. For example, in one of the data sets, individuals are on average more impatient when they are asked about two sub-intervals – today to 6 months, and 6 months to 12 months – than when they are asked to make choices between today and 12 months. This same pattern is replicated across various time horizons, different stake sizes, and both data sets: Individuals tend on average to be more impatient for short than long time horizons, but similarly impatient regardless of when the short time horizon starts. The same pattern is observed for different sub-populations defined based on gender, age, education, and financial situation, so it appears to be a general feature of intertemporal choice. Robustness checks yield no evidence that the pattern is explainable by credibility issues, concavity of utility, arbitrage strategies, or non-stationary preferences and credit

[5] Harrison et al. (2002) were 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 OH type designs. Read et al. (2005) studied sub-additivity in a relatively large, but non-representative, internet experiment, in which 1 out of about 2,000 participants was randomly selected to receive real money.

constraints. An individual level analysis confirms the aggregate results and shows the pervasiveness of additivity violations: Roughly 70 percent of individuals violate additivity, with almost all of these being sub-additive. The prevalence of non-additive choices is not attenuated by individual level characteristics, except that an indicator of financial sophistication (thinking about market interest rates) is associated with a reduced propensity to violate additivity.

One implication of the paper is to underline the value of non-traditional approaches to modeling inter-temporal choice, which can incorporate the sub-additive choice pattern.[6] The results do not rule out that individuals are "present biased," potentially in addition to being sensitive to framing effects.[7] There are various other types of evidence for present bias and self control problems, for example from the field.[8] Our results show, however, an important way in which present-bias is not a complete description of how time horizon variation affects inter-temporal choice.[9]

Another implication of the paper is methodological. The findings suggest that the typical approach of using a single type of time horizon comparison to identify discounting types can be misleading in a systematic way. For example, when we look at OH comparisons in our data, we replicate the typical findings from previous studies using this design, i.e., on average individuals appear hyperbolic (roughly 60 percent at the individual level).[10] With our SH comparisons, by contrast, we would conclude that the same group of subjects exhibits constant discounting on average, but with fully mixed results at the individual level in the form of roughly equal proportions constant, declining, and increas-

[6] Various models have been proposed than might explain Read's findings of sub-additivity, including Rubinstein (2003), Scholten and Read (2006), Ebert and Prelec (2007), Zauberman et al. (2009), and Epper and Fehr-Duda (2012).

[7] See Laibson, 1997 and O'Donoghue and Rabin, 1999 for seminal papers on present bias.

[8] See, e.g., Read and Van Leeuwen, 1998; Della Vigna and Malmendier, 2006; Ariely and Wertenbroch, 2002; Ashraf et al., 2006; Milkman et al., 2009; Augenblick et al., 2015. Results less consistent with present-bias are found by Harrison et al., 2005; Read et al., 2012; and Halevy, 2015.

[9] Various studies have documented individual heterogeneity in quantitative estimates of the rate of time discounting, assuming exponential discounting and using consumption data (e.g., Lawrance, 1991; Cagetti, 2003; Alan and Browning, 2010). Our results show substantial individual heterogeneity in intertemporal choice patterns, but in qualitative as well as quantitative terms.

[10] Frederick et al. (2002) survey ten early studies that use an OH approach and all find behavior consistent with declining discounting. Frederick et al. also note that the same pattern arises when one compares across roughly thirty studies that use either short or long time horizons. The survey also includes four studies that use as SH approach, which find results that could be consistent with declining discounting, but some of these studies have some unusual design features that make interpretation more difficult. Bisin et al. (2010) is an example of a more recent study using an OH design and finding evidence that appears consistent with declining discounting.

ing discounting patterns. This is in line with more mixed results typically found in the literature with this design.[11] A third type of comparison that is possible in our data would yield increasing discounting on average for the same group of people (roughly 60 percent at the individual level). These contrasting results map, broadly speaking, onto the pattern of mixed findings in the literature, and follow directly from the sub-additive choice pattern: Greater impatience for short than long horizons implies seemingly hyperbolic behavior in an OH comparisons, whereas similar impatience for short horizons regardless of timing will lead to the appearance of constant discounting on average in SH comparisons. Our findings suggest a value in using multiple types of time horizon comparisons for the same individual to help avoid potential misclassification. Based on our findings, this is likely to reveal that individuals who would otherwise be classified as exponential or hyperbolic are in fact making choices that are inconsistent with any standard discounting model.

The rest of the paper is organized as follows. Section 2 describes the data sets, treatments, and behavioral predictions of standard discounting models. Section 3 presents the aggregate results, and Section 4 provides an analysis of individual heterogeneity. Section 5 discusses potential policy implications and concludes.

2 Design of the Experiment

2.1 Data Collection and Experimental Procedures

Our analysis uses two data sets. One data set, which we call the SOEP data involves a 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 Schupp and Wagner, 2002; Wagner et al., 2007). The second data set is the SOEP Cross Sectional Study, which we call CSS data for short. This data set involves a separate sample of individuals, collected by the SOEP administration on an annual basis as part of the process of "pretesting" questions for potential use in the SOEP survey.

Data collection for both data sets was done by the same professional surveying company that administers the SOEP every year, using the same sampling procedure as for

[11] See, e.g., Kirby and Herrnstein, 1995; Anderhub et al. (2001); Andreoni and Sprenger (2012); and Sutter et al. (2013).

the SOEP.[12] Both the CSS and SOEP samples were constructed so as to be representative of the adult population, age 17 and older, living in Germany.[13] Subjects were visited by interviewers in their own homes. In total the CSS data include 500 subjects who participated in the intertemporal choice experiments, and the SOEP data include 1,503 such subjects.

Participants in our studies went through a computer-assisted personal interview (CAPI) conducted with a laptop. First, subjects answered a detailed questionnaire. Topics included demographic characteristics, financial situation, health, and attitudes. The full questionnaire, in German and translated into English, is available upon request. Second, at the end of the questionnaire, subjects were invited to participate in a paid experiment.

The first step in the experimental procedure involved the experimenter presenting subjects with some example choices, explaining the types of choices the subject would face, and how payment would work. Once there were no more questions, the experiment began. An example of the script and instructions used in the experiments is presented in Appendix A.4 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 T_{st} is defined by starting date s, and ending date t. For a given horizon an individual chose between an early payment, X, available at the start of the horizon, and series a larger, later payments, Z, available at the end of the horizon (in an abuse of notation, in this section we suppress superscripts for early and late payments indicating time horizon). For example, in horizon T_{012}, the early payment is at time 0, and the later payment is 12 months in the future. In all choices for a given horizon, the amount of X was held constant, but the later payment, Z, was larger in each subsequent choice. For most time horizons, the value of Z 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 implied an additional 2.5 percentage point increase in the annual rate of return, up to a maximum of 50 percent.[14] In one treatment, we implemented choices with a coarser

[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.

[14] We chose semi-annual compounding of the annual interest rate because this is a natural compromise

measurement of the IRR, in steps of 5 percentage points, but in this case measured annual IRRs as high as 100 percent. We obtain a measure of the IRR needed to induce an individual to wait for a given time horizon by observing the smallest value of Z that induces them to wait. More precisely, we obtain upper and lower bounds for the annual IRR, separated by 2.5 percentage points, due to the discrete variation in late payment amounts; in our analysis, we focus on lower bounds. Across treatments, we varied s and t as well as the amounts X and Z.

IRRs were elicited in an incentive compatible manner: Subjects were presented with the different choices, one at a time, on the computer screen, and subjects knew that one choice would be randomly selected at the end of the experiment and implemented. Furthermore, subjects knew that at the end of the experiment a random device would determine whether they were actually paid, with the probability of being paid equal to 1/7 in the CSS and 1/9 in the SOEP data. This procedure gave subjects an incentive to choose according to their true preferences in each choice situation.

A key feature of our design was the high credibility of payments. One feature that contributed to credibility of payments for both data sets was the fact that the agency conducting the experiments is well-known and trusted by the German public.[15] Interviewers also left their contact details at the end of the experiment, making it easy for subjects to contact the institute, but there were no reports, from any of the interviewers, about subjects expressing concerns regarding credibility of payments. There is an even stronger argument for credibility in the case of our SOEP data, however, because all participants were members of the SOEP panel itself. Unlike participants in most intertemporal choice experiments, these individuals were in a long-term relationship with the individual surveyor conducting the experiment. Thus, payments should have been highly credible regardless of timing. We did also use a "front-end delay" approach to achieving equal credibility of early and late payments that has often been used in the literature (see Coller and Williams, 1999). Specifically, all payments arrived by mail after

---

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.

[15] The agency is employed as a reliable and reputable polling company by German public television stations; television new programs regularly feature results from the agency's opinion polls on social and political issues.

the experiment. Thus, early payments did not have any special credibility arising from being paid during the experiment session itself.

A further robustness check regarding credibility is possible because we varied the payment procedure across data sets. Specifically, for the CSS data set, all payments were mailed on the day after the interview and thus would arrive within at most two days due to the well-known two-day guarantee for delivery by the German postal service. Checks for immediate payments could be cashed immediately, once they arrived, while checks for payments in the future were post-dated through a special arrangement with the issuing bank and could only be cashed at the specified time. For the SOEP data, by contrast, payments were also sent by mail, but the timing of the mailing reflected the timing of the payments, i.e., checks for immediate payments were mailed immediately after the experiment and arrived within two days, while checks for later payments were mailed punctually on the corresponding later date. In this case neither check was post-dated. If credibility concerns are an important issue for subjects, one would expect to see that changes in the payment procedure affect results. In our analysis we compare results across the two data sets.

A final note on the design concerns the length of the front-end delay. Many previous studies that tested discounting assumptions have featured 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, such a delay might tend to reduce "immediacy" of early payments. This could matter for the predictions of some non-constant discounting models, which assume a discrete drop in discount rates between the present and the future. The verdict is still out on how quickly such a drop might occur, but if it is assumed to occur between the present and the next day, then both early and late payments in our experiments are beyond the time frame in which present bias exerts its influence. In such a case, the model would make still make testable predictions, but these would be the same as those of the constant discounting model, as discussed below. Making a trade-off between credibility and immediacy, we chose the shortest possible front-end delay compatible with avoiding a same-day credibility problem.[16]

[16] Mainly for the purposes of concreteness, so as to avoid having to say 1 or 2 days repeatedly, the experimental instructions told subjects that immediate rewards would be referred to as being received "Today"[quotes included]. At the same time, the instructions were very clear that all rewards would arrive after the experiment by post, and that: "Today means you can cash the check you receive by post immediately." Ultimately, we see little evidence that this wording lead to differential behavior in

Table 1: Summary of Treatments

Measure	Data set	Sub-sample	Early payment (in Euro)	Upper-bound IRR	Obs.
T_{012}	CSS	n.a.	100	52.5%	500
T_{06}	CSS	n.a.	100	52.5%	500
T_{612}	CSS	n.a.	100	52.5%	500
T_{012}	SOEP	1 & 2	200	52.5%	977
T_{06}	SOEP	1	200	52.5%	490
T_{01}	SOEP	2	200	52.5%	487
T_{01b}	SOEP	3	200	105%	526
T_{1213}	SOEP	3	200	105%	526

2.2 Treatments

Table 1 summarizes the various treatments. As shown in the table, the CSS data involved three different measures of annual IRR for each subject: 0 to 6 months (T_{06}), 0 to 12 months (T_{012}), and 6 months to 12 months (T_{612}). The order of the treatments in the CSS data was randomized across individuals. The early payment was always 100 euros, and the largest delayed payment always implied an annual IRR of 50 percent (compounded semi-annually) for waiting the specified length of time. If individuals never chose the later payment, their IRR was right-censored, and coded as having a (lower-bound) value of 52.5 percent.

In the SOEP data we used higher stakes, different treatments, and different payment procedures, in order to assess whether the results from CSS replicate with different parameter values and design features. The early payment was always 200 euros in the SOEP and thus stakes were higher than in the CSS, even accounting for the lower payment probability of 1/9 rather than 1/7.[17] The SOEP also differed from the CSS in having only two treatments for each individual (see Table 1). This was due to time constraints in the survey implementation. We varied these treatments across sub-samples, however, in a way that complements the CSS study design and led to a total of five different treatments.

Specifically, in the SOEP data, the first sub-sample of 490 individuals had measures

early versus later time horizons. For example, observed impatience is similar regardless of whether the time horizon involves early payments "Today" or involves early payments in 12 months.

[17] 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.

for 0 to 6 months (T_{06}) and 0 to 12 months (T_{012}). This allows an OH comparison with time horizon lengths that are directly comparable to the CSS data. The second sub-sample of 487 were asked about 0 to 1 month (T_{01}) and 0 to 12 months (T_{012}), shedding light on how OH comparisons change as the discrepancy in horizon length increases. For the third sub-sample of 526 we measured discounting for 0 to 1 month (T_{01b}) and 12 to 13 months (T_{1213}), to allow an SH comparison, with a relatively large time delay between horizons. The measures for the third sub-sample were also different because IRRs were measured in steps of 5 percent rather than 2.5 percent, and the upper-bound IRR in each horizon was 105 percent rather than 52.5 percent. This was designed to help explore the nature of discounting without potential censoring at 52.5 percent. We denote the one-month measure in this sub-sample T_{01b}, to distinguish it from T_{01} in the second sub-sample.

Order was predetermined in the SOEP data: for the first two sub-samples, the T_{012} measure was always elicited first; for the third sub-sample, T_{01b} was elicited first. A random device on the computer selected whether an individual was assigned to the first, second, or third sub-sample experiments.

2.3 Behavioral Predictions

In this section we present predictions derived from different traditional models of time discounting – constant and non-constant – as well as predictions based on the sub-additivity hypothesis. To derive the predictions we employ assumptions that are often maintained, explicitly or implicitly, in the literature using intertemporal choice experiments. These include workhorse assumptions like time separability, but also the following: (1) equal credibility of payments regardless of timing or amount; (2) utility is locally linear; (3) people treat monetary payments like consumption opportunities and do not engage in arbitrage between the interest rates offered in the experiment and market interest rates; (4) a time-stationary period utility function that does not vary with calendar date. The recent literature has raised concerns about assumptions (1) to (4), so later in the analysis we discuss whether our results might be driven by the failure of one of these assumptions.

We illustrate the predictions by considering, without loss of generality, an example with three time horizons, T_{06}, T_{012}, and T_{612}, corresponding to the structure of the CSS data. The early payment is always 100 for each horizon, compounding occurs once every 6 months, and for simplicity a period is assumed to be 6-months long. We assume for now that early payments at time 0 are literally available on the day of the experiment, when preferences are measured.

When making decisions in T_{06}, T_{012}, and T_{612}, 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:

$$T_{06}:\left(1+\frac{r}{2}\right)100 \lesseqgtr \mathbf{Z^{T_{06}}};\quad T_{012}:\left(1+\frac{r}{2}\right)^2 100 \lesseqgtr \mathbf{Z^{T_{012}}};\quad T_{612}:\left(1+\frac{r}{2}\right)100 \lesseqgtr \mathbf{Z^{T_{612}}}$$

where $\mathbf{Z^{T_{06}}}$, $\mathbf{Z^{T_{012}}}$, and $\mathbf{Z^{T_{612}}}$ denote sets of later payments available for the corresponding time horizons. We denote by $Z^{T_{06}}$, $Z^{T_{012}}$, and $Z^{T_{612}}$ the lowest element of these sets of later payments, which make the individual indifferent between the earlier or the later payments for the corresponding time horizon. Observations on $Z^{T_{06}}$, $Z^{T_{012}}$, and $Z^{T_{612}}$ obtained from the experiments establish the points of indifference for each horizon, and define the internal rates of return:

$$\left(1+\frac{IRR^{T_{06}}}{2}\right)100 = Z^{T_{06}};\quad \left(1+\frac{IRR^{T_{012}}}{2}\right)^2 100 = Z^{T_{012}};\quad \left(1+\frac{IRR^{T_{612}}}{2}\right)100 = Z^{T_{612}} \tag{1}$$

We neglect the fact that the delayed payment is actually a variable measured on a discrete grid in the experiment, and thus that we can infer only a range for the IRR; this has no consequences for the qualitative predictions, and eases exposition.[18] Solving for IRRs yields:

$$IRR^{T_{06}} = 2\left(\frac{Z^{T_{06}}}{100}-1\right);\quad IRR^{T_{012}} = 2\left(\left(\frac{Z^{T_{012}}}{100}\right)^{\frac{1}{2}}-1\right);\quad IRR^{T_{612}} = 2\left(\frac{Z^{T_{612}}}{100}-1\right) \tag{2}$$

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

2.3.1 Constant (Exponential) Discounting

In the case of constant discounting, an individual is indifferent between the early and delayed payments in T_{06}, T_{012}, and T_{612} when

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

$$\left(1+\frac{\rho}{2}\right)100 = Z^{T_{06}}; \qquad \left(1+\frac{\rho}{2}\right)^2 100 = Z^{T_{012}}; \qquad \left(1+\frac{\rho}{2}\right)100 = Z^{T_{612}}, \tag{3}$$

where ρ is the constant rate of time preference. As this is the same as condition (1), it follows directly that a constant discounter chooses $Z^{T_{06}}$, $Z^{T_{012}}$, and $Z^{T_{612}}$ such that the measured IRR is invariant with respect to time horizon: $IRR^{T_{06}} = IRR^{T_{012}} = IRR^{T_{612}} = \rho$.

2.3.2 Declining and Increasing Discounting

In the case of declining discounting, e.g., as in the case of a hyperbolic discount function, there are different *six-month* discount rates, ρ_1 and ρ_2, for periods 1 and 2 (first 6 months and second 6 months), such that $\rho_1 > \rho_2$. In this case the points of indifference in T_{06}, T_{012}, and T_{612} are given by

$$(1+\rho_1)\,100 = Z^{T_{06}}; \qquad (1+\rho_1)\,(1+\rho_2)100 = Z^{T_{012}}; \qquad (1+\rho_2)\,100 = Z^{T_{612}}. \tag{4}$$

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

$$IRR^{T_{06}} = 2[(1+\rho_1)-1]; \quad IRR^{T_{012}} = 2[((1+\rho_1)(1+\rho_2))^{\frac{1}{2}}-1]; \quad IRR^{T_{612}} = 2[(1+\rho_2)-1]. \tag{5}$$

Given $\rho_1 > \rho_2$, this implies $IRR^{T_{06}} > IRR^{T_{012}} > IRR^{T_{612}}$. Intuitively, impatience should be greatest in T_{06} with declining discounting, because it includes the present and extends the least far into the future. Behavior in T_{612} should be the most patient, because it excludes the present and only concerns payments relatively far into the future. Behavior in T_{012} should be in-between. An analogous argument establishes that with increasing discounting, the model predicts the opposite ranking of IRRs by time horizon, $IRR^{T_{06}} < IRR^{T_{012}} < IRR^{T_{612}}$.

The commonly-used approximation to the hyperbolic discounting model, the quasi-hyperbolic or β - δ model (e.g., Phelps and Pollak, 1968; Laibson, 1997; O'Donoghue and Rabin, 1999), assumes a discrete drop in the discount rate between the present and the near future, and then constant discount rates between all more-distant future periods. In the literature studies differ in their assumptions about how quickly after the present the drop in discount rates occurs. If the present bias is assumed to extend more than

2 days but less than 6 months into the future, then the quasi-hyperbolic model makes the same predictions as models with hyperbolic, or other forms, of continuously declining discounting, $IRR^{T_{06}} > IRR^{T_{012}} > IRR^{T_{612}}$. If the drop in discount rates is assumed to occur within the two-day window between the date of the experiment and the arrival of the early payments, then only the constant discounting part of the quasi-hyperbolic discount function is relevant for the choices in our experiment, and the model predicts $IRR^{T_{06}} = IRR^{T_{012}} = IRR^{T_{612}}$, the same as with constant discounting. 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). We show in the Appendix that while this model can predict a "magnitude effect,", i.e., higher measured patience as stake sizes increase, for a given level of stakes it generates the same qualitative predictions for IRRs across time horizons as the declining discounting or quasi-hyperbolic models.

2.4 Sub-additivity

The sub-additivity hypothesis is that individuals are less patient when a time horizon is framed in sub-intervals. By contrast, constant and non-constant discounting models predict that this change in framing has no effect on measured IRR's.

To see this, note that in our example choices in T_{012} reveal a delayed payment, $Z^{T_{012}}$, sufficient to induce waiting. In the framework of the constant and non-constant discounting models, this reflects compounding of two six-month discount rates, ρ_1 and ρ_2, with $\rho_1 = \rho_2$ in the case of constant discounting or $\rho_1 \neq \rho_2$ in the case of non-constant discounting. From (2) we have:

$$Z^{T_{012}} = 100\left(1 + \frac{IRR^{T_{012}}}{2}\right)^2 = 100(1+\rho_1)(1+\rho_2). \tag{6}$$

In the same framework we should be able to take the IRRs observed for T_{06} and T_{612} and construct a late payment $Z^{T'_{012}}$ that should make the individual indifferent between 100 at time 0 or waiting 12 months to get $Z^{T'_{012}}$. From (2) we have:

$$Z^{T'_{012}} = 100\left(1 + \frac{IRR^{T_{06}}}{2}\right)\left(1 + \frac{IRR^{T_{612}}}{2}\right) = 100(1+\tilde{\rho}_1)(1+\tilde{\rho}_2). \tag{7}$$

In the case of the standard discounting models, it must be the case that $\rho_1 = \tilde{\rho_1}$ and $\rho_2 = \tilde{\rho_2}$, since we are eliciting patience over the exact same periods regardless of whether we ask about sub-intervals or the whole interval. In the appendix we show that this must also hold for any arbitrary discount function, since if it is not true that $\rho_1 = \tilde{\rho_1}$ and $\rho_2 = \tilde{\rho_2}$, one can construct choices that will lead the individual to violate transitivity.

If framing matters for intertemporal choice, however, the two delayed payments, and implied six-month discount rates, may differ. Furthermore, if the result is sub-additivity, the answers will differ such that the sub-intervals imply a larger delayed payment needed to induce waiting, i.e., $Z^{T'_{012}} > Z^{T_{012}}$. This in turn implies

$$100\left(1+\frac{IRR^{T_{012}}}{2}\right)^2 < 100\left(1+\frac{IRR^{T_{06}}}{2}\right)\left(1+\frac{IRR^{T_{612}}}{2}\right) \quad (8)$$

Or equivalently, $(IRR^{T_{012}})^2 < IRR^{T_{06}} \cdot IRR^{T_{612}}$.

3 Aggregate Results on IRR and Time Horizon

In this section we present results on the aggregate intertemporal choice patterns, and then discuss these in light of the behavioral predictions.

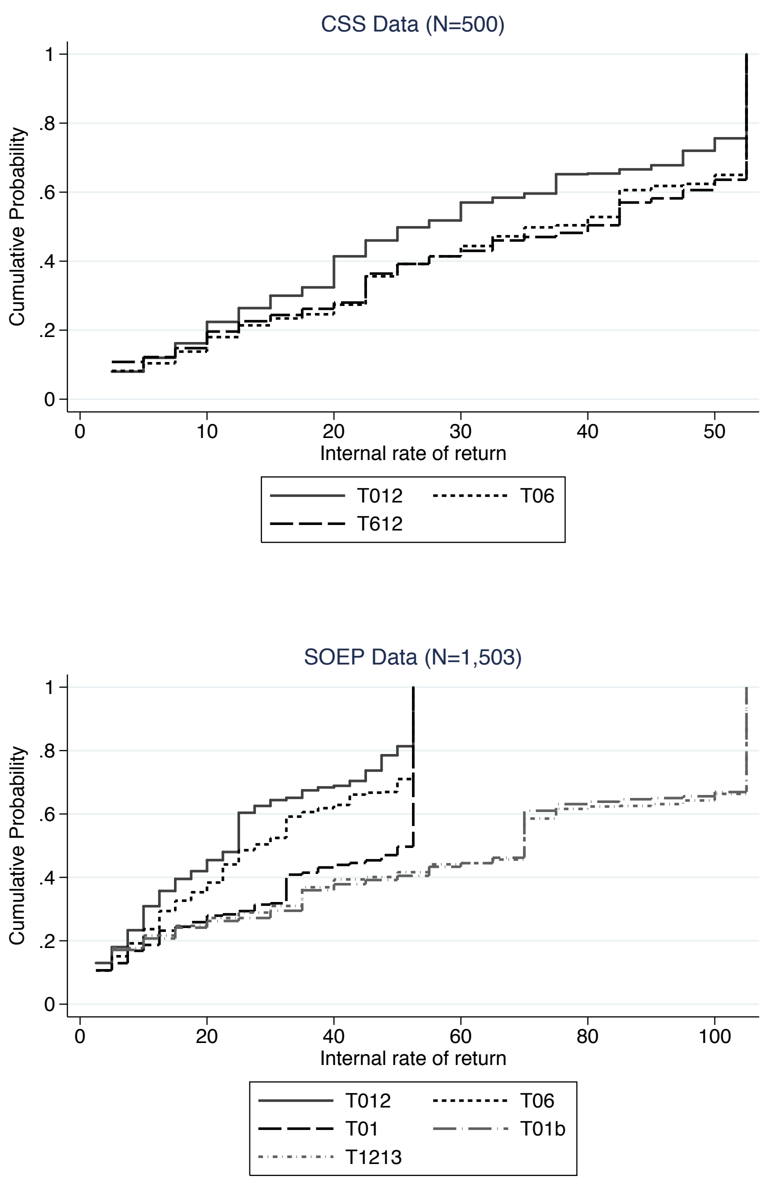
3.1 Comparisons of Cumulative Distributions

Figure 1 presents the cumulative distribution functions of the annual IRR for each of the different time horizons. The top panel shows results from the CSS data, and the bottom panel shows results from the SOEP data. For the SOEP data, we pool the T_{012} measures across the two sub-samples that have this treatment, as the order and stake sizes are identical.[19]

In the top panel of Figure 1, we see that people in the CSS data exhibit a different distribution of IRRs for T_{012} than for T_{06}, or T_{612} ($p < 0.001$; $p < 0.001$; Kolmogorov-Smirnov), in the direction of lower IRRs and greater patience for the 12-month horizon. People are quite similar in impatience, however, comparing the distributions of IRRs for the T_{06} and T_{612} measures, and there is no significant difference ($p < 0.90$; Kolmogorov-Smirnov). Thus, people tend to be more patient for long horizons than short horizons, but similarly impatient over the short horizons regardless of the starting date. The quantitative magnitudes of the difference in IRRs is also substantial: The median IRRs for T_{06} and T_{612} are 10 percentage points higher than the median IRR for T_{012}.

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

Figure 1: Cumulative Distributions of IRR



The bottom panel shows the same patterns in the SOEP data. IRRs increase monotonically as time horizon length decreases, with greater impatience for T_{06} than T_{012} ($p < 0.001$; Kolmogorov-Smirnov), and even greater impatience for T_{01} than T_{06} ($p < 0.001$; Kolmogorov-Smirnov). At the same time, IRRs are relatively insensitive to the starting date of the time horizon, in that people are similarly impatient for T01b and T1213 ($p < 0.98$; Kolmogorov-Smirnov). Again, differences are large: The IRR for T_{01} is more than 25 percentage points higher than the median IRR for T_{012}. The distribution for T_{01} is almost identical to T_{01b} and T_{1213}, except for a deviation in the direction of greater patience starting around the middle of the range for T_{01}. This is potentially due to either an order effect, or a framing effect of the different upper bound for the IRR, or both.[20] Regardless, the cumulatives for the one-month horizons are more similar to each other than to the cumulatives for longer horizons.

3.2 Interval Regression Analysis

Tables 2 and Table 3 provide another way to look at the results, 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, clustering on individual.

Column (1) of Table 2 shows that IRRs in the CSS data are significantly lower for the T_{012} measure compared to T_{06}, by more than 6 percentage points, while there is not a significant difference between T_{612} and T_{06}.[21] Table 3 presents similar regression analysis based on the SOEP data. Results are reported separately for the three sub-samples, in Panel A, B, and C, respectively. Looking at Column (1) we again see a pattern of lower IRRs for longer horizons, but similar IRRs across horizons of the same length regardless of the starting date: T_{012} is significantly lower than T_{06} by about 5 percentage points, and lower than T_{01} by about 20 percentage points, while T_{1213} is not significantly different from T_{01b} and the point estimate indicates only a 1 percentage point difference.

[20] Over the common support, the distributions for T_{01b} and T_{1213} are each significantly different from the distribution for T_{01} ($p < 0.001$; $p < 0.001$; Kolmogorov-Smirnov).

[21] The regression analysis also allows checking robustness with respect to order effects, in the CSS data where order was randomized. We add dummy variables for the different possible treatment orders, and interactions of T_{012} and T_{612} with all of the different orders, to the specification used in Column (1) of Table 2. We again find a significant difference between T_{012} and T_{06}, but not between T_{06} and T_{612}. Furthermore, all interaction terms are not statistically significant, individually or jointly.

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

	All	Males	Females	Age ≤med.	Age >med.	Less educated	More educated	Income ≤med.	Income >med.	Credit Constrained	Not Credit Constrained	More risk Averse	Less risk Averse	Did not Consider Market Interest	Considered Market Interest
	(1)	(2)	(3)	(4)	(5)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
$T012$	-6.42***	-8.20***	-4.82***	-6.56***	-6.34***	-6.63***	-5.83***	-5.75***	-6.85***	-5.03**	-6.64***	-6.07***	-6.66***	-8.55***	-3.93***
	(0.85)	(1.20)	(1.19)	(1.05)	(1.39)	(1.04)	(1.42)	(1.32)	(1.10)	(2.17)	(0.92)	(1.12)	(1.19)	(1.20)	(1.22)
$T612$	0.14	0.25	0.07	-1.26	1.75	0.71	-1.30	2.05	-1.11	3.80	-0.77	1.44	-0.66	-1.80	2.82*
	(1.11)	(1.50)	(1.63)	(1.47)	(1.70)	(1.32)	(2.08)	(1.72)	(1.45)	(2.91)	(1.21)	(1.40)	(1.59)	(1.63)	(1.46)
Constant	37.22***	35.96***	38.33***	35.76***	39.14***	39.21***	31.84***	39.08***	35.90***	42.32***	34.97***	36.08***	37.94***	46.84***	23.83***
	(1.40)	(1.98)	(1.99)	(1.71)	(2.34)	(1.69)	(2.45)	(2.08)	(1.87)	(3.11)	(1.56)	(2.21)	(1.82)	(2.07)	(1.62)
Observations	1500	693	807	768	732	1113	387	591	909	330	1131	564	936	936	558

Notes: Interval regression estimates. Dependent variable is the IRR for a given time horizon. The reference time horizon is T_{06}. Age and household income groups are defined relative to median values. More educated indicates that an individual completed the Abitur, a college entrance exam in Germany. Credit constraints are measured by a question asking about ability to borrow money in the event of an unexpected expense. In parentheses, robust s.e., adjusted for clustering on individual. *, ** indicates significance at 10 and 5 percent level.

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

	All (1)	Males (2)	Females (3)	Age ≤med. (4)	Age >med. (5)	Less educated (8)	More educated (9)	Income ≤med. (10)	Income >med. (11)	Credit Constrained (12)	Not Credit Constrained (13)	More risk Averse (14)	Less risk Averse (15)	Did not Consider Market Interest (16)	Considered Market Interest (17)
							Panel A								
Sample 1:															
T012	-5.26***	-4.86***	-5.66***	-5.90***	-4.58***	-5.33***	-4.97***	-3.20**	-5.77***	-5.07***	-5.33***	-5.03***	-5.40***	-6.61***	-3.58***
	(0.68)	(0.89)	(1.03)	(0.94)	(0.98)	(0.77)	(1.44)	(1.49)	(0.76)	(1.87)	(0.72)	(0.91)	(0.97)	(0.95)	(0.97)
Constant	31.17***	28.31***	34.15***	30.48***	31.85***	34.11***	21.46***	29.91***	31.49***	43.85***	27.63***	34.83***	28.31***	37.99***	21.45***
	(1.34)	(1.86)	(1.91)	(1.87)	(1.91)	(1.62)	(2.17)	(2.98)	(1.49)	(3.22)	(1.41)	(1.97)	(1.79)	(1.94)	(1.57)
Observations	980	504	476	486	494	730	218	198	782	230	744	418	562	622	358
							Panel B								
Sample 2:															
T012	-19.67***	-21.05***	-18.46***	-19.10***	-20.28***	-20.15***	-16.64***	-19.90***	-19.60***	-5.07***	-5.33***	-5.03***	-5.40***	-6.61***	-3.58***
	(1.35)	(2.04)	(1.80)	(1.97)	(1.86)	(1.62)	(2.39)	(2.64)	(1.57)	(1.87)	(0.72)	(0.91)	(0.97)	(0.95)	(0.97)
Constant	44.96***	46.61***	43.51***	41.08***	48.84***	47.50***	32.35***	42.77***	45.80***	43.85***	27.63***	34.83***	28.31***	37.99***	21.45***
	(1.80)	(2.64)	(2.47)	(2.42)	(2.70)	(2.11)	(3.51)	(3.18)	(2.18)	(3.22)	(1.41)	(1.97)	(1.79)	(1.94)	(1.57)
Observations	974	444	530	468	506	766	180	238	736	230	744	418	562	622	358
							Panel C								
Sample 3:															
T1213	-1.45	-0.22	-2.42	-0.10	-3.14	1.78	-7.48	-6.56	0.23	-3.60	-0.93	1.77	-4.50	-7.05	5.04
	(3.36)	(5.03)	(4.52)	(4.33)	(5.27)	(4.15)	(6.03)	(6.94)	(3.84)	(7.12)	(3.70)	(4.81)	(4.70)	(4.95)	(4.22)
Constant	97.25***	96.53***	97.83***	91.92***	103.49***	102.48***	82.34***	96.81***	97.37***	139.69***	89.92***	102.17***	92.05***	122.69***	57.85***
	(5.16)	(7.72)	(6.94)	(6.79)	(7.89)	(6.36)	(8.85)	(10.48)	(5.93)	(18.21)	(5.25)	(7.46)	(7.11)	(7.77)	(5.52)
Observations	1052	464	588	552	500	784	240	254	798	188	860	544	508	722	330

Notes: Interval regression estimates, separately by sub-sample of the SOEP data. Dependent variable is the IRR for a given time horizon, with two time horizons (observations) per individual. Reference category is T_{06} in Panel A, T_{01} in Panel B, and T_{01b} in Panel C. Age and household income groups are defined relative to median values. More educated indicates that an individual completed the Abitur, a college entrance exam in Germany. Credit constraints are measured by a question asking about ability to borrow money in the event of an unexpected expense. In parentheses, robust s.e., adjusted for clustering on individual. *, ** indicates significance at 10 and 5 percent level.

In columns (2) through (11), Table 2 and Table 3 explore whether similar results are observed for different sub-groups in the population. We consider sub-populations defined by gender, age, education level, and income level. Looking at the regression estimates, it is apparent that the qualitative results on time horizon effects are similar across all of the sub-groups, in both data sets. Thus, the patterns we find are not isolated to specific populations, but rather are a general feature of intertemporal choice.

3.3 Discussion: Aggregate Results and Model Predictions

To summarize, the aggregate results reveal a consistent choice pattern with two key features: (1) People are more impatient for short than long time horizons; (2) people are relatively insensitive to when a given short time horizon starts. In the CSS data this is shown by $IRR^{T_{06}} = IRR^{T_{612}} > IRR^{T_{012}}$, and in the SOEP data, by $IRR^{T_{01b}} = IRR^{T_{1213}} \approx IRR^{T_{01}} > IRR^{T_{06}} > IRR^{T_{012}}$.[22]

These findings are not well explained by either constant, declining, or increasing discounting, maintaining the usual identifying assumptions. The sensitivity of IRRs to time horizon length is inconsistent with constant discounting, while the insensitivity of IRRs to starting date of a given horizon is contrary to the key prediction of declining or increasing discounting (and also to recent two-system models that predict declining discounting; see, e.g., Fudenberg and Levine, 2012). The data are also not consistent with alternative versions of the declining discounting model, quasi-hyperbolic or fixed cost, regardless of assumptions about the length of the present.

The aggregate results are consistent with the sub-additivity hypothesis. The patterns we observe imply that, on average, $(IRR^{T_{012}})^2 < IRR^{T_{06}} \cdot IRR^{T_{612}}$. Thus, impatience over 12 months is greater when it is elicited in two sub-intervals, than when it is elicited directly for the entire 12 month horizon.

The aggregate results also show that different measurement approaches tend to lead

[22] It is noteworthy that IRRs tend to be lower in the SOEP data than the CSS data, for comparable time horizons (e.g., compare the constant term in Column (1) of Table 2 to the constant term in Column (1), Panel A, of Table 3). One interpretation is a "magnitude effect", such that the level of impatience decreases with stake size for a given pair of horizons. Such an effect would be an anomaly for standard discounting models, where the IRR is assumed to be independent of stake size. Many other studies also report finding lower IRRs as stake sizes increase (see Frederick et al., 2002). One explanation is provided by the model of fixed-cost declining discounting, proposed by Benhabib et al., 2010. An alternative explanation for this stylized fact could be that the unobserved utility function is more linear for higher stakes (see Andersen et al., 2011, for a discussion).

to systematically different conclusions, when considered in isolation. In the CSS data, the OH comparison of today to 6 months with today to 12 months would yield the conclusion that the average individual exhibits declining discounting, replicating the typical finding of studies using this design. Using the SH comparison of today to 6 months vs. 6 months to 12 months, however, one would conclude that the same group of people is consistent with constant discounting, on average. A third type of comparison that is possible in the CSS data involves comparing discounting over a long and short time horizon ending on the same date, e.g., today to 12 months vs. 6 months to 12 months; we denote this an Overlapping Shifted Horizon (OSH) comparison. Based on the OSH comparison in the CSS data, we would conclude that the typical individual is an increasing discounter. The same pattern of method variance in results arises in the SOEP data. The two OH comparisons, used in isolation, would lead to the conclusion that the average individual is a declining discounter, whereas the SH comparison yields choices that look like constant discounting on average. These seemingly contradictory results follow directly from how different types of measures capture different portions of the sub-additive choice pattern: Because IRR's are higher for short than long horizons, results from OH comparisons will look like declining discounting, and results from OSH comparisons will look like increasing discounting; similar IRR's for short horizons, regardless of start date, will lead to the appearance of constant discounting in SH comparisons.

3.4 Investigating Sensitivity to Maintained Assumptions

The theoretical predictions were based on maintained Assumptions (1) to (4), which have been questioned in the recent literature. In this section we discuss whether failures of one or more of these assumptions could be driving the pattern of aggregate results. Columns (12) to (17) of Table 2 and Table 3 present some results that are useful as robustness checks.

Assumption (1) is that the payments in the experiment were not differentially credible across different time horizons. If payments are not perfectly credible, and credibility varies with both timing of the payment and payment amount, then particular patterns of credibility concerns could generate the sub-additive choice pattern we observe.[23] As

[23] One can construct examples in which this is the case, involving credibility that declines with distance of the promised payment date from the present, and letting perceived credibility vary with the payment

discussed above, however, there are several features of our setting that help ensure that credibility of all payments is very high regardless of timing, for example subjects being in a long-term relationship with the experimenter. Furthermore, we varied the payment procedure across data sets, in a way that might be expected to generate different results if subjects have credibility concerns. Instead, we have now seen the same qualitative choice patterns in both datasets. For these reasons, we conclude that credibility concerns are unlikely to be a key mechanism driving the results.

If utility is concave rather than approximately linear for the range of stakes in the experiment, this violates Assumption (2), and could generate choices that are qualitatively in line with the sub-additivity findings. Intuitively, diminishing marginal utility can have a stronger impact on longer time horizons than shorter ones, because by construction, longer time horizons involve larger delayed payments.[24]

To explain our results, however, concavity would need to be sufficiently extreme to generate the observed quantitative magnitudes of differences in choices. Assuming a standard CRRA utility function, one would need a coefficient of relative risk aversion equal to 35 to equalize the true IRR's for short and long horizons.[25] This is the same magnitude as the coefficient value that would be needed to explain the equity premium puzzle, and is

amount. Somewhat counter-intuitively, these examples involve credibility increasing with the payment amount.

[24] If the estimates of impatience are affected by concavity of the unobserved utility function, $u(\cdot)$ with $u'(\cdot) > 0, u''(\cdot) < 0$, we have $\widetilde{IRR}^{T_{06}} = 2(\frac{u(Z^{T_{06}})}{u(100)} - 1)$, $\widetilde{IRR}^{T_{012}} = 2((\frac{u(Z^{T_{012}})}{u(100)})^{\frac{1}{2}} - 1)$, and $\widetilde{IRR}^{T_{06}} = 2(\frac{u(Z^{T_{612}})}{u(100)} - 1)$. By concavity, it is clear that $\widetilde{IRR}^{T_{06}} < IRR^{T_{06}}$, $\widetilde{IRR}^{T_{012}} < IRR^{T_{012}}$, and $\widetilde{IRR}^{T_{612}} < IRR^{T_{612}}$. For our purposes, however, the more relevant question is how the *relative* sizes of IRRs for different time horizons change, allowing for concavity. There are two effects that may work in opposite directions. On the one hand, introducing concavity leads to a bigger reduction in the value of the ratio $Z^{Tst}/100$ for the longer time horizon, e.g., $Z^{T_{012}}/100 - u(Z^{T_{012}})/u(100) > Z^{T_{06}}/100 - u(Z^{T_{06}})/u(100)$. This works in the direction of reducing $IRR^{T_{012}}$ by more than the IRRs for short horizons, implying that we underestimate the true discrepancy between IRRs for short and long horizons by assuming linearity. On the other hand, a given change in the ratio in late payment to early payment causes a larger decrease in IRRs for short horizons than long horizons, due to compounding (this is immediately observable from the relevant derivatives of IRRs with respect to these ratios). Thus, whether or not concavity understates or exaggerates the difference in IRRs depends on the relative sizes of these different effects.

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

$$\frac{\widetilde{IRR}^{T_{012}}}{\widetilde{IRR}^{T_{06}}} = \frac{\left(\frac{Z^{T_{012}}}{100}\right)^{1-\theta} - 1}{\left(\frac{Z^{T_{06}}}{100}\right)^{1-\theta} - 1}; \qquad \frac{\widetilde{IRR}^{T_{012}}}{\widetilde{IRR}^{T_{612}}} = \frac{\left(\frac{Z^{T_{012}}}{100}\right)^{1-\theta} - 1}{\left(\frac{Z^{T_{612}}}{100}\right)^{1-\theta} - 1} \tag{9}$$

Using $\theta = 35$, the exponential discounting model can explain the different observed median delayed payments for T_{06}, T_{012}, and T_{612}: 119, 129, and 120 Euros, respectively.

typically regarded as implausibly high (see, e.g., Siegel and Thaler, 1997). Furthermore, recent empirical estimates of θ, from lottery experiments with similar stake sizes to our intertemporal choice experiments, are far smaller.[26]

Another approach to evaluating the role of concavity is to check whether the results are related to an indicator for the degree of risk aversion. In our data we have a survey measure of risk aversion, which has been shown in previous research to correlate with measures of risk aversion based on real stakes lottery experiments (Dohmen et al., 2011).[27] Using the survey measure, Columns (14) and (15) of Table 2 compare the results across the sub-groups with less than or greater than the median willingness to take risks. A very similar sub-additivity pattern is observed for both groups. This provides another indication that concavity of utility is unlikely to drive the sub-additivity result.

While the literature has typically assumed that subjects treat monetary rewards in discounting experiments as equivalent to consumption opportunities, it could be that assumption (3) is not valid, and subjects engage in arbitrage between the experiment and market interest rates. Although it is not obvious what would be the arbitrage opportunities that would generate the sub-additive choice pattern, such opportunities are typically unobserved by the researcher, leaving open the possibility that these could play a role. In our design, however, we included a measure of whether individuals might be thinking about arbitrage: After subjects had completed the experiment we asked whether they had thought about market interest rates at all during the experiment, a seeming pre-requisite for engaging in arbitrage between experiment and market interest rates. The majority of subjects in both data sets stated that they had not not done so (67% in the CSS data, and 66% in the SOEP data). Columns (14) and (15) show that the sub-additivity pattern is also present for individuals who did not think about arbitrage possibilities. This indicates that the sub-additivity result is not driven by arbitrage strategies.

[26] The median parameter value is estimated to be in the range of 0.55 to 0.75 (Andersen et al., 2008; Andersen et al., 2010). In this approach, lottery experiments generate modest variations in earnings, and risk aversion is assumed to reflect diminishing marginal utility of money over this range. Other estimates, based on large-stakes (hypothetical) lottery questions administered in a large representative survey, show parameters in a similar range, and suggest that $\theta = 35$ is completely outside the support of the empirical distribution (see Dohmen et al., 2005).

[27] The question asks: "How willing are you, in general, to take risks." Respondents can use an 11-point scale from completely unwilling (0) to completely willing (10). For a representative sample of German adults who responded to the the question, and also participated in real stakes lottery experiments, answers to the question were robustly correlated with the degree of risk aversion measured in the experiments (Dohmen et al., 2011).

Assumption (4) is that preferences are stationary over time. Halevey (2015) argues that one source of non-stationarity in preferences could be the possibility of randomly arising expenses, combined with credit constraints. More generally, non-stationarity could arise if for some reason people have different preferences at different calendar dates. Columns (16) and (17) of both tables show that the sub-additivity pattern is similar regardless of whether or not individuals report being credit constrained, indicating that this source of non-stationarity does not appear to drive the results. We have also seen similar patterns regardless of calendar dates, e.g., discounting is similar across short time horizons regardless of whether they are 6 months apart (today vs. 6 and 6 vs. 12) or one year apart (today vs. 1 month vs. 12 months vs. 13 months). This provides additional evidence that the results are not driven by preferences changing with calendar date.

4 Individual Level Results

A first observation about individual-level heterogeneity in our data is that OH, SH, and OSH approaches, if used in isolation, would lead to very different distributions of "discounting types." This mirrors the method variance discussed above for results at the aggregate level. Specifically, in Table A.1 in the Appendix, we show the following: The modal discounting pattern for OH comparisons is declining discounting (roughly 60 percent); there is a roughly uniform distribution across declining, constant, and increasing types for SH measures; the OSH approach has a modal pattern consistent with increasing discounting (roughly 60 percent).

To gain a more holistic picture of an individual's type of intertemporal choice pattern we combine all the measures we have for the same person, focusing on the CSS data because there we have all three types of comparisons (OH, SH, and OSH) for each person. Table 4 reports fractions of individuals falling into "additive," "sub-additive", and "super-additive" categories. The categorization takes into account that the experiments measure IRRs in intervals of 2.5 percentage points, and is deliberately conservative in terms of favoring additivity: Someone is classified as sub-additive if the squared upper-bound of the IRR for the long time horizon is lower than the product of the lower bounds of the short horizon IRRs, and vice versa for super-additivity. The residual group is classified as additive. The analysis excludes individuals for whom right-censoring makes the classification

Table 4: Individual Types, CSS Data

Types	Additive	Sub-additive	Super-additive
Percent	28.45	60.22	11.33
		$N = 362$	
Types, allowing for larger "errors" (favors additivity)	**Additive**	**Sub-additive**	**Super-additive**
Percent	48.67	43.66	7.67
		$N = 339$	

Notes: The first row shows the percentages of individuals in the CSS data who exhibit additivity, sub-additivity, and super-additivity, excluding individuals for whom right-censoring prevents unambiguous classification. In the second row we adopt a more conservative approach to classifying individuals as violating additivity, which favors additivity.

ambiguous.[28]

Looking at the first row of Table 4, we see that about 28 percent of individuals satisfy additivity, e.g., by having the same IRR for all three horizons. Thus, the group of individuals whose choices are consistent with one of the standard discounting model is a minority. By contrast, about 60 percent of individuals exhibit sub-additivity. The remaining 11 are super-additive. Thus, sub-additivity is by far the most prevalent choice pattern at the individual level.

In additional analysis we explored whether additivity violations might be driven by people who have exactly the same IRR interval for the two short time horizons, which is different from their IRR interval for the long horizon. This might conceivably reflect a choice heuristic, such that individuals choose the same IRR for both of the short time horizons because they involve exactly the same monetary amounts (an unavoidable feature of SH comparisons), neglecting to pay attention to the difference in timings. It turns out, however, that the majority is still sub-additive if we exclude such individuals from the sample. Indeed, more than half of those who are sub-additive do not have exactly the same IRR interval for the two short horizons.

[28] This construction of the sample takes into account the fact that right-censoring of short time horizons does not prevent classifying someone as sub-additive, and right censoring of the long horizon does not prevent categorizing someone as super-additive. The sample involves $N = 362$.

We also checked robustness of additivity violations to an even more conservative approach that further expands the upper and lower bounds for each time horizon's IRR, effectively allowing for some larger "errors" in decision making, and favoring additivity. Specifically, we added and subtracted 2.5 percentage points from the regular upper and lower bounds, respectively. We then classified someone as sub-additive only if the square of the wider upper bound IRR for the long horizon is less than the product of the wider lower bound IRR's for the short horizons, and vice versa for super-additive. Again, we exclude individuals for whom right-censoring makes classification ambiguous.[29] The second row of Table 4 shows the results. While this approach mechanically favors additivity, we see that a narrow majority still violates additivity, and sub-additivity remains a key feature of the distribution of types. Specifically, 51.33 percent of individuals still violate additivity, and 44 percent are sub-additive.

As a last step, we exploit our representative data to explore the relationship of additivity violations to a range of characteristics. Table 5 presents the results of Probit estimations, where the dependent variable is equal to 1 if an individual violates additivity and 0 otherwise. The sample excludes individuals for whom right-censoring leads to an ambiguity in assessing additivity violations. The independent variables include the same demographics used for the sub-group analysis in the section on aggregate results, but we also include a quadratic in age, and a measure of cognitive ability based on the combined score on two short-form tests.[30]

The results in Table 5 show that additivity violations are not significantly more or less pronounced based on variation in the different demographic traits, including cognitive ability. An exception is thinking about market interest rates: Those who thought about market interest rates are less likely to violate additivity. This goes in the opposite direction of arbitrage strategies leading to apparent additivity violations. One explanation could be that it captures a type of financial sophistication or numeracy that helps people avoid

[29] Note that the combination of allowing for errors, but addressing the impact of censoring on categorization, leads to a smaller sample. For example, consider individuals who were unambiguously sub-additive without errors, but had right censoring for one of the short time horizons. Allowing for errors, some of these are no longer unambiguously sub-additive, but on the other hand, right-censoring prevents classifying them as ambiguously additive. The resulting sample involves $N = 339$.

[30] The scores on the tests are standardized and then averaged to form a single measure of cognitive ability. One test involved matching numbers and unfamiliar symbols for 90 seconds, capturing speed of processing, and the other involved naming as many animals as possible in 90 seconds, providing a measure of crystalized intelligence. Both tests correlate with corresponding sub-modules of widely used tests of cognitive ability. For a more detailed description of these tests see Dohmen et al. (2010).

additivity violations.[31] The results reinforce the conclusions from sub-group analysis at the aggregate level, that additivity violations are prevalent across a wide range of groups defined by demographic characteristics or financial situation.

Table 5: Correlates of non-additive discounting at the individual level

	Dep. variable: =1 if non-additive			
	(1)	(2)	(3)	(4)
Female	0.04	0.04	0.03	-0.06
	(0.14)	(0.15)	(0.15)	(0.17)
Age	-0.02	-0.03	-0.02	0.01
	(0.02)	(0.02)	(0.02)	(0.02)
Age Squared	0.00	0.00	0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Cognitive ability		-0.03	0.01	0.05
		(0.10)	(0.11)	(0.11)
More educated			-0.25	-0.15
			(0.17)	(0.18)
Ln(household income)				-0.20
				(0.14)
Willingness to take risks				-0.01
				(0.03)
Thought about int. rate				-0.64***
				(0.17)
Credit constrained				0.08
				(0.22)
Constant	1.01**	1.12**	1.11**	2.33**
	(0.45)	(0.49)	(0.49)	(1.17)
Observations	361	319	319	300

Notes: Probit estimates, with robust standard errors in parentheses. Sample excludes individuals for whom right-censoring prevents assessing additivity violations.

5 Conclusion

This paper provides new evidence that sub-additivity is a robust and pervasive feature of intertemporal choice. The design addresses potentially important confounds, which have been raised about intertemporal choice experiments with time-dated monetary rewards, and shows that these do not drive the sub-additive choice pattern. The findings have implications for economic theory, underling the value of research on models that can incorporate sub-additivity. They also raise caveats about the ways that time horizon effects in intertemporal choice experiments have traditionally been interpreted. This latter implication may help explain mixed results from previous studies, attempting to measure

[31] A multinomial logit that treats sub-additivity and super-additivity as separate categories yields a similar conclusion, showing that individuals who thought about market interest rates were significantly less likely to be sub-additive or super-additive.

the prevalence of discounting types, or using experiment-based measures of discounting types to explain economic outcomes.

A final point concerns the potential policy relevance of the findings. One implication concerns the on-going debate about policy interventions designed to address self-control problems (e.g., Camerer et al., 2003). Such policies are often motivated with reference to models with declining (hyperbolic or quasi-hyperbolic) discounting, and the impact of such policies depends crucially on the distribution of discounting types in the population. Intertemporal choice experiments have been a commonly used tool for measuring the prevalence of declining discounting. Our results raise caveats, however, about the typical experimental design for measuring the distribution of types, which has used choices over two time horizons, elicited in the present, to try to trace out the shape of a time discount function. The observed impact of time horizon on choices may reflect a mechanism that is that is outside of the hyperbolic or any other standard discounting model. Instead, combining multiple types of time horizon comparisons, or using alternative types of experiments, may be better suited to identify present bias. A second policy-related implication of the paper is that a different bias in decision making, sub-additivity, is also worth consideration by policy makers it its own right, as it affects many people in the population. Sub-additivity implies that framing time horizons in more narrow or more broad ways my influence how people make choices. This can be relevant when policy makers design and advertise different types of packages of benefits or services to be received or payed for over time, and potentially for assessing the welfare effects of different ways that firms might describe choices to consumers.

References

Alan, S., and M. Browning (2010): "Estimating intertemporal allocation parameters using synthetic residual estimation," *Review of Economic Studies*, 77(4), 1231–1261.

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

Andersen, S., G. Harrison, M. Lau, and E. Rutstroem (2010): "Discounting behavior: A reconsideration," *European Economic Review*, 71, 15–33.

——— (2011): "Discounting Behavior and the Magnitude Effect: Evidence from and Field Experiment in Denmark," *Economica*, 80(320), 670–697.

Andreoni, J., and C. Sprenger (2012): "Estimating Time Preferences from Convex Budgets," *American Economic Review*, 102(7), 3333–3356.

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

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

Augenblick, N., M. Niederle, and C. Sprenger (2015): "Working over time: Dynamic inconsistency in real effort tasks," *Quarterly Journal of Economics*, pp. 1067–1115.

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 (2012): "Which measures of time preference best predict outcomes: Evidence from a large-scale field experiment," *Journal of Economic Behavior and Organization*, 84(1), 308–320.

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

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

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

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

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

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

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

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

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

Epper, T., and H. Fehr-Duda (2015): "The missing link: Unifying risk taking and time discounting," Discussion paper, Working Paper Series, Department of Economics, University of Zurich.

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

Halevy, Y. (2015): "Time consistency: Stationarity and time invariance," *Econometrica*, 83(1), 335–352.

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

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

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

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

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

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

Martinez, Seung-Keun and Meier, Stephan and Sprenger, Charles (2017): "Procrastination in the Field: Evidence from Tax Filing," *UC San Diego Working Paper*.

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

Meier, S., and C. D. Sprenger (2013): "Discounting financial literacy: Time preferences and participation in financial education programs," *Journal of Economic Behavior and Organization*, 95, 159–174.

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

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," *Review of Economic Studies*, 35(2), 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," *working paper, London School of Economics and Political Science*.

Read, D., S. Frederick, and M. Airoldi (2012): "Four days later in Cincinnati: Longitudinal tests of hyperbolic discounting," *Acta Psychologica*, 140(2), 177–185.

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

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

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), 1424–1436.

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

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

Sutter, M., M. G. Kocher, D. Glätzle-Rützler, and S. T. Trautmann (2013): "Impatience and uncertainty: Experimental decisions predict adolescents' field behavior," *American Economic Review*, 103(1), 510–531.

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

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*.

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.

A. Appendix

A.1 Additional Tables

Table A.1: Individual Types by Measurement Approach

Measurement approach	Time horizons	Sample	Constant	Declining	Increasing
				Percent	
OH	T_{06} vs. T_{012}	CSS	22.39	60.70	16.92
SH	T_{06} vs. T_{612}	CSS	27.89	34.37	37.75
OSH	T_{012} vs. T_{612}	CSS	15.92	19.15	64.93
OH	T_{06} vs. T_{012}	SOEP	25.92	54.28	19.80
OH	T_{01} vs. T_{012}	SOEP	12.97	76.06	10.97
SH	T_{01b} vs. T_{1213}	SOEP	32.91	32.91	34.18

A.2 Predictions for Fixed-Cost Specification of the Quasi-Hyperbolic Model

The fixed-cost version of the quasi-hyperbolic model allows this model to predict a "magnitude effect," a tendency for measured impatience to decrease as stake sizes increase, *ceteris paribus*. Besides predicting a magnitude effect, however, this model has the same properties as the quasi-hyperbolic model: There is an extra cost of waiting, above and beyond exponential discounting, when comparing the present to the next future period. When comparing two adjacent, future periods, however, the extra cost applies to both periods, and thus cancels out, leaving discounting to be governed solely by the exponential discount rate. The quasi-hyperbolic model with fixed costs thus predicts $IRR^{T_{06}} > IRR^{T_{012}} > IRR^{T_{612}}$ if the present extends further than two days into the future, and invariance of the IRR to time horizon, $IRR^{T_{06}} = IRR^{T_{012}} = IRR^{T_{612}}$, if it does not, the same as the standard quasi-hyperbolic model.

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

$$\left(1+\tfrac{\rho}{2}\right)100 + \left(1+\tfrac{\rho}{2}\right)b = Z^{T_{06}} \quad \left(1+\tfrac{\rho}{2}\right)^2 100 + \left(1+\tfrac{\rho}{2}\right)^2 b = Z^{T_{012}}$$

$$\left(1+\tfrac{\rho}{2}\right)100 = Z^{T_{612}}$$

Substituting into (2) yields

$$IRR^{T_{06}} = 2\left[\left(1+\tfrac{b}{100}\right)\left(1+\tfrac{\rho}{2}\right) - 1\right] \quad IRR^{T_{012}} = 2\left[\left(1+\tfrac{b}{100}\right)^{\frac{1}{2}}\left(1+\tfrac{\rho}{2}\right) - 1\right]$$

$$IRR^{T_{612}} = 2\left[\left(1+\tfrac{\rho}{2}\right) - 1\right]$$

These equations imply $IRR^{T_{06}} > IRR^{T_{012}} > IRR^{T_{612}}$. 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, $IRR^{T_{06}} = IRR^{T_{012}} = IRR^{T_{612}}$. 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.

A.3 Demonstration of transitivity violation

To see more formally why the aggregate choice pattern in the CSS data implies a violation of transitivity, note that T_{06}, T_{012}, and T_{612} pose individuals with choices between different bundles that involve payments received at 0, 6, or 12 months. In the data, we observe for each time horizon the late payment amounts $Z^{T_{06}}$, $Z^{T_{012}}$, and $Z^{T_{612}}$ such that an individual is indifferent between the bundle involving the early payment and the bundle involving the late payment. The table below shows these particular bundles for each horizon:

	T_{06}		T_{012}		T_{612}
$A:$	$(100, 0, 0)$	$A:$	$(100, 0, 0)$	$D:$	$(0, 100, 0)$
$B:$	$(0, Z^{T_{06}}, 0)$	$C:$	$(0, 0, Z^{T_{012}})$	$E:$	$(0, 0, Z^{T_{612}})$

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

An inconsistency arises, however, when one compares across choice pairs. From T_{012}, we observe the $Z^{T_{012}}$ that makes an individual indifferent between receiving 100 at time 0 or $Z^{T_{012}}$ in 12 months

$$Z^{T_{012}} = 100\left(1 + \frac{IRR^{T_{012}}}{2}\right)\left(1 + \frac{IRR^{T_{012}}}{2}\right) = 100(1+\rho_1)(1+\rho_2). \tag{A.1}$$

Combining the observed IRRs for T_{06} and T_{612}, we can also construct a late payment $Z^{T'_{012}}$ that should make the individual indifferent between 100 at time 0 or waiting 12 months to get $Z^{T'_{012}}$

$$Z^{T'_{012}} = 100\left(1 + \frac{IRR^{T_{06}}}{2}\right)\left(1 + \frac{IRR^{T_{612}}}{2}\right) = 100(1+\tilde{\rho}_1)(1+\tilde{\rho}_2). \tag{A.2}$$

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

$$\left(\frac{1}{1+\rho_1}\right)\left(\frac{1}{1+\rho_2}\right) \neq \left(\frac{1}{1+\tilde{\rho}_1}\right)\left(\frac{1}{1+\tilde{\rho}_2}\right). \tag{A.3}$$

I.e., total discounting over 12 months is different, based on the 6-month measures, than it is based on the 12-month measure. This immediately implies (from (A.1) and (A.2)) that $Z^{T_{012}} \neq Z^{T'_{012}}$.

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

	T'_{012}
A	$(100, 0, 0)$
C'	$(0, 0, Z^{T'_{012}})$

and since $Z^{T_{012}} \neq Z^{T'_{012}}$, should not be indifferent for $Z^{T_{012}}$. I.e., $A \sim C' \implies A \nsim C$. But combining these implied preferences with the preferences that are actually revealed in T_{012} for late payment $Z^{T_{012}}$, we have $A \sim C \nsim C' \sim A$. This implies $A \sim C \nsim A$, a violation of transitivity.

There is also a violation of transitivity relaxing the assumption of linear utility, and allowing for a concave (monotonic) utility function $u(\cdot)$, as long as (A.3) still holds. As discussed above, (A.3) holds even if utility is concave, unless the utility function is characterized by implausibly extreme concavity. With non-linear utility the late payments satisfy the following conditions:

$$u\left(Z^{T_{012}}\right) = u(100)(1+\rho_1)(1+\rho_2); \qquad u\left(Z^{T'_{012}}\right) = u(100)(1+\tilde{\rho}_1)(1+\tilde{\rho}_2). \qquad \text{(A.4)}$$

.

If it is still the case that $(1+\rho_1)(1+\rho_2) \neq (1+\tilde{\rho}_1)(1+\tilde{\rho}_2)$ after allowing for concavity, we clearly have $Z^{T_{012}} \neq Z^{T'_{012}}$, and there is an inconsistency in preferences that again implies a violation of transitivity.

A.4 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. Can 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.