STABILITY AND ADAPTIVITY: PREFERENCES OVER TIME AND UNDER RISK

ISBN: 978 90 3610 468 5

This book is no. 681 of the Tinbergen Institute Research Series, established through cooperation between Rozenberg Publishers and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Cover illustration by Delfina De Lillo.

Stability and Adaptivity: Preferences over time and under risk

Stabiliteit en adaptiviteit:
besluitvorming over tijd en onder risico

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. H.A.P. Pols

and in accordance with the decision of the Doctorate Board

The public defense shall be held on
Friday 3 February 2017 at 09.30 hours
by

YU GAO
born in Taiyuan, China.

Erasmus University Rotterdam

Doctoral Committee

Promotors:	Prof.dr. H. Bleichrodt
	Prof.dr. K.I.M. Rohde
Other Members:	Prof.dr. A. Baillon
	Prof.dr. O. l'Haridon
	Prof.dr. G. van de Kuilen

謹以此書献給我的父母

Acknowledgements

There are so many people to thank for helping me during my PhD. The four years wouldn't be so enjoyable without you.

Firstly, I would like to express my sincere gratitude to Prof. Han Bleichrodt and Prof. Peter Wakker for bringing me to their amazing research group, and their continuous support of my Ph.D. study with their patience, motivation, encouragement and immense knowledge. Their guidance helped me in all the time of research and writing of this thesis. I could not imagine having better mentors for my Ph.D. study.

Besides, I thank Prof. Kirsten Rohde for her long time support and guidance with expertise which dates back to the writing of my master thesis. She not only gave me advice to academic work, but also suggestions which are especially valuable to junior researchers. My sincere thanks also goes to Prof. Aurelien Baillon, for his insightful comments, patience and inspiration. With his influence I grabbed the tools of lab experiment programming. This skill helped my complete involvement in every experiment of the research projects and deepened my understanding of the work of an experimenter.

I thank my coauthors Arthur Attema, Ilke Aydogan, Zhenxing Huang and Chen Li for the stimulating discussions, for the precious days we were working together, for the things I learned from you, and for your tolerance of my stubbornness. I am so lucky to work with you.

I thank the classmates in Tinbergen Institute and fellows in Erasmus. I thank Li Wei for his help and unselfishness since our master study. I thank Dai Meimei for the joys we had together and the aspirations we shared with each other. I thank Wang Tong for all the "addictive" adventures we had. You know I mean reading papers! I thank Zheng Jindi for the enjoyable discussions on research ideas and all the roller coaster rides in Gardaland. I will surely remember that day! I thank Yang Jingni for her kindness and trust. I wish we had more time to hang around. My special thanks goes to my dear officemate Hale Koc. We were raised in two geographically and culturally remote countries but we have so much in common. I enjoyed every day, including all the silence, and all the short and long talks with you in our office.

I thank Rogier Potter van Loon, Dennie van Dolder, Vitalie Spinu, Jan Stoop, Uyanga Turmunkh for all the years we spent in the same research group, for the discussions on Thursday meetings, for the dinner parties we had, and for the games we played together. I thank Ge Jing for her accompany although we never worked together, I could not imagine my Ph.D. without you. In addition, I thank Jelmer van der Gaast for his kindness and generous help.

Last but not least, my deep gratitude goes to my best friend and life partner Ning. Thank you for coming in my life. You have made it a dream come true.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

People's preferences have some stable features. When making decisions involving time, we are usually impatient: we seek for immediate gratification, and care less about future outcomes. Sometimes, we do not appreciate the urge to pursue immediate gratification. We make plans for the future, but disappointedly, fail to stick to them when the time is approaching. There are always things that we know we should do, but never do. It appears that our impatience is not constant, but stronger when the time is nearer: impatience declines over time.

In decisions under risk, we tend to overweight events that happen with very small probabilities, and underweight events that happen with intermediate and large probabilities. We also have stronger preference for avoiding losses to acquiring gains. This tendency of loss aversion is not just for huge stakes like life or your house, but also for stakes as small as 10 dollars. Imagine a person is going to toss a fair coin, if it is tails, you lose $10. How much would you have to gain on winning in order for this gamble to be acceptable to you? A lot of people will ask for more than $20.

Preferences are also adaptable. Although in general we prefer sooner rewards than later, the degree to which delayed rewards are discounted, and the impatience changes, varies across domains. Imagine losing weight in a healthy way, we are recommended to do both physical exercise and dieting. For some people, it is easier to stick to gym plans than to dieting; for others, it is the opposite. It seems that our mental strength and self-control do not have the same power for every aspect in life, but differ across domains. Similarly, decisions under risk are susceptible to context. The same person could prefer "In a group of 600 people, 200 people will be saved" over "1/3 probability of saving all 600 people" and at the same time prefer "2/3 probability that 600 people will all die" over "In a group of 600 people, 400 will die", despite that the two sets of choices are essentially the same.

This dissertation addresses both the stability side, and the adaptivity side of decisions over time and under risk. Because of the stable features of preferences, we could use the same models and functions to describe preferences in different domains and under different context. The first two chapters are dealing with the measurements of time discounting and risk attitudes. Also, because preferences are adaptable, we should bear in mind that conclusions on one kind of outcomes or from one specific context cannot be directly applied to all outcomes or scenarios. In the latter three chapters we explore the richness of preference adaptivity.

Chapter 2: How to measure discounting reliably and easily?

As the most widely used model to analyze intertemporal choices, Discounted Utility evaluates future outcomes by their utility weighted by a discount factor. The existing methods to measure discounting either make the questionable assumption that utility in risky choices is the same as utility in riskless intertemporal choices, or use complex methods to elicit utility and discounting at the same time, and are therefore susceptible to collinearity between utility and discounting.

This chapter presents a tractable method to measure discounting that requires no knowledge of utilities, because they cancel out. The cancelling out of utilities requires a critical assumption: time separability, which means that what you had in the past and what you will have in the future do not affect the utility of what you have now. Since this assumption is violated sometimes (imagine that the utility of your lunch might be affected by what you ate in the morning and what you will eat for dinner), we tested it in the experiment, and found that separability holds for most people under our context. We also compared our results with a traditional, utility based method introduced by Epper et al. (2011). We found that our method needs fewer questions but gives similar results.

Chapter 3: Is Prelec's probability weighting function reliable?

Prelec's (1998) compound-invariant family provides an appealing way to model probability weighting and is widely used in empirical studies. Prelec (1998) gave a behavioral foundation for this function, however, this condition is hard to test empirically as it requires a lot of indifference. Luce (2001) proposed a simpler condition: reduction invariance, which is easier to test empirically.

In this chapter, reduction invariance is tested in a lab experiment. The data support reduction invariance both at aggregate level and at individual level where reduction invariance was the dominant pattern. The descriptive validity of the compound-invariant weighting function is confirmed. In latter chapters, we use Prelec's function to analyze risky choices.

Chapter 4: Do people discount health and money differently?

Both individuals and governments make decisions involving outcomes that occur at different points in time. Examples are choosing a pension plan, purchasing durable goods and funding scientific projects. Constant discounting is usually assumed, due to its traceability and normative appeal, although it is not frequently observed in reality. Loosely speaking, a person who discounts constantly will get up in the morning when the alarm clock set by herself wakes her up, and goes to gym at the frequency she planned when paying for the membership. In addition, it is still unclear whether health and money should be discounted similarly. Policy implications require more accurate descriptions on people's time preferences.

This chapter measures deviations from constant discounting for health and money. Our method allows to compare the degree of decreasing impatience between the two domains. The results indicate that most subjects deviated from constant discounting and were decreasingly impatient for both money and health. Further, this deviation is larger for health than for money.

Chapter 5: Does money make you irrational?

Money plays a significant role in people's lives. Our social preferences and individual behaviors change when we are around money. Vohs and Schooler (2008) has shown that people who are exposed to money suddenly become less helpful than those who aren't. Kouchaki et al. (2013) showed that people are more likely to lie or make immoral decisions after being exposed to money-related words. Prelec and Simester (2001) found that consumers' willingness to pay decreases substantially when using cash instead of credit cards. In this chapter, we investigate if holding cash influences people's risk attitudes.

In an experiment, we studied simple lottery valuation tasks, and implemented two treatments: number and cash. In the number treatment, the outcomes of lotteries were denoted by numbers. In the cash treatment, the lottery outcomes were presented with real cash and subjects were asked to value the lotteries with real notes and coins. Subjects in the cash treatment gave lower valuations than those in the number treatment: they were more risk averse. By fitting our data to Prelec's probability weighting function, we conclude that cash makes people less sensitive to probability changes, but has no effect on pessimism.

Chapter 6: Are black swans really ignored?

There are two kinds of uncertainties in life. For the first type, you have information sources like weather forecasts, drug-package inserts, and mutual fund brochures, all of which provide descriptions of possible outcomes (rainy or sunny, various complications, potential profits) and probabilities. For the second, you have no summary descriptions of possible outcomes or their likelihoods, such as whether to go out on a date, when to pass a truck on the highroad, or to take part in dangerous sports or not. For the later events, you have to rely on your own encounters with different occasions, and make decisions from experience (DFE).

Hertwig et al. (2004) and a lot of studies after them found that DFE and decisions from descriptions (DFD) can lead to dramatically different choice behaviors. In particular, under DFE, people seem to ignore events that happen with small probabilities, which is opposite to DFD. This chapter explores the DFE-DFD gap by resolving problems in former studies, which enables us to observe the genuine weightings of probabilities. Overall, our findings suggest a clear de-biasing effect of sampling experience: it attenuates - rather than reverses - the commonly found inverse-S shaped probability weighting in DFD.

Chapter 2

Measuring Discounting without Measuring Utility

Discounted utility is the most widely used model to analyze intertemporal decisions. It evaluates future outcomes by their utility weighted by a discount factor. Measuring discount factors is difficult because they interact with utility. Most measurements simply assume that utility is linear [1], which is unsatisfactory for many economic applications. Frederick et al. (2002, p.382), therefore, suggested measuring utility through risky choices while assuming expected utility, as in Chapman (1996b), and then to use these utilities to measure discount factors. In the health domain, this method had been used before for flow (continuous) variables (Stiggelbout et al., 1994). In economics, Andersen et al. (2008) and Takeuchi (2011) used this method for discrete outcomes.

The aforementioned method has two limitations. First, expected utility is often violated (Starmer, 2000), which distorts utility measurements. Second, the transfer of risky cardinal utility to riskless intertemporal choice is controversial (Raiffa and Luce, 1957, p.32 Fallacy 3; Camerer, 1995, p.619; Moscati,2013;). Andreoni and Sprenger (2012b) and Abdellaoui et al. (2013) provided empirical evidence against such a trans-

This chapter is based on the homonymous paper, co-authored with Arthur Attema, Han Bleichrodt, Zhenxing Huang and Peter P. Wakker. Online appendix can be accessed through: https://assets.aeaweb.org/assets/production/files/1466.pdf

[1] See Warner and Pleeter (2001), Frederick et al. (2002, p.381),Tanaka et al. (2010), Sutter et al. (2013).

fer. When introducing discounted utility, Samuelson (1937; last paragraph) immediately warned that cardinal intertemporal utility may differ from other kinds of cardinal utility. To avoid these two difficulties, some studies elicited both utility and discounting from intertemporal choices (Abdellaoui et al., 2010; Andreoni and Sprenger, 2012a,b; Abdellaoui et al., 2013; Epper and Fehr-Duda, 2015). Such elicitations are complex and susceptible to collinearities between utility and discounting.

This paper presents a tractable method to measure discounting that requires no knowledge of utility. We adapt a recently introduced method for flow variables in health (Attema et al., 2012) to discrete monetary outcomes in economics. Flow variables, such as quality of life, are continuous in time and are consumed per time unit. Whereas theoretical economic studies sometimes take money as a flow variable, experimental studies of discounting invariably take it as discrete, received at discrete timepoints, and we will do so here. Because our method directly measures discounting, and utility plays no role, we call it the direct method (DM).

The basic idea of the DM is as follows. Assume that a decision maker is indifferent between: (a) an extra payment of \$10 per week during weeks 1-30; and (b) the same extra payment during weeks 31-65. Then the total discount weight of weeks 1-30 is equal to that of weeks 31-65. We can derive the entire discount function from such equalities. Knowledge of utility is not required because it drops from the equations. Even though this method is elementary, it has not been known before.

The DM is easy to implement and subjects can easily understand it. In an experiment, we compare it with a traditional, utility based, method (UM). In our comparison, we use the implementation of the UM by Epper et al. (2011, EFB henceforth), which is based on prospect theory, currently the most accurate descriptive theory of risky choice. We show that the DM needs fewer questions than the UM but gives similar results.

2.1 Theory

We assume a preference relation $\succcurlyeq$ over discrete *outcome streams* $(x_1,...,x_T)$, yielding *outcome* (money amount) x_j at time t_j, for each $j \leq T$. For ease of presentation, we consider the stimuli used in our experiment, where $T = 52$ and the unit of time is one week. Thus $(x_1,...,x_{52})$ yields x_j at the end of week j, for each j. *Discounted utility* holds if preferences maximize the *discounted utility* of outcome stream x:

$$\sum_{j=1}^{52} d_j U(x_j) \tag{2.1}$$

Here, U is the subjective *utility function*, which is strictly increasing and satisfies $U(0) = 0$, and $0 < d_j$ is the subjective *discount factor* of week j. For $E \subset \{1, ..., 52\}$, $\alpha_E\beta$ denotes the outcome stream that gives outcome α at all timepoints in E and outcome β at all other timepoints. $C(E)$ denotes the cumulative sum $\sum_{j \in E} d_j$ and reflects the total time weight of E. $C(k)$ denotes C ({ $1, ..., k$ }). C is called the *cumulative (discount) weight* . The proof of the following result clarifies why we do not require utility: it drops from the equations.

OBSERVATION 1. Assume discounted utility, and $\alpha > \beta$. Then:

$$\alpha_A\beta \succ \alpha_B\beta \Leftrightarrow \sum_{j \in A} d_j > \sum_{j \in B} d_j \quad (i.e., C(A) > C(B)); \tag{2.2}$$

$$\alpha_A\beta \sim \alpha_B\beta \Leftrightarrow \sum_{j \in A} d_j = \sum_{j \in B} d_j \quad (i.e., C(A) = C(B)); \tag{2.3}$$

$$\alpha_A\beta \prec \alpha_B\beta \Leftrightarrow \sum_{j \in A} d_j < \sum_{j \in B} d_j \quad (i.e., C(A) < C(B)); \tag{2.4}$$

PROOF. The preference and two inequalities in Eq.2.2 are each equivalent to $C(A)(U(\alpha) - U(\beta)) + C(\{1, ..., 52\})U(\beta) > C(B)(U(\alpha) - U(\beta)) + C(\{1, ..., 52\})U(\beta)$. The other results follow from similar derivations.

Using Observation 1, we can derive equalities of sums of d_js, which, in turn, define the function C on $\{1, ..., 52\}$ and all the d_js. This procedure does not require any knowledge of utility and is therefore called the *direct method (DM).*

In the mathematical analysis, we also consider a continuous extension of C, defined on all of $(0, 52]$, and also called the cumulative (discount) weight. At the timepoints $1, ..., 52$ it agrees with C defined above. In the continuous extension, any payoff x_j is a salary received during week j. Receiving a salary of x_j per week during week j amounts to receiving x_j at time j. We equate j with $(j-1, j]$ here. Salary can also be received during part of a week. In the continuous extension, $C(t)U(\alpha)$ is the subjective value of receiving α during period $(0, t]$, where $C(t) = C(0, t]$ and t may be a noninteger, $0 \leq t \leq T$. Then $C(t, 52] = C(0, 52] - C(0, t]$ also for nonintegers t. In all the empirical estimations reported later, we extend C from integers to nonintegers using linear interpolations. Given the small time interval of a week, a piecewise linear approximation is satisfactory. The following remark shows that the d_js serve as discretized approximations of the derivative of C.

REMARK 2. $d(j) = C(j) - C(j-1)$ is the average of the derivative C' over the interval $(j-1, j]$. Thus, d_j is approximately $C'(t)$ at $t = j$.

2.2 Measuring discounting using the Direct Method

We now explain how C can be measured up to any degree of precision using the DM. Of course, $C(0) = 0$. We normalize $C(52) = d_1 + ... + d_{52} = 1$. We write $c_p = C^{-1}(p)$. Then $c_0 = 0$ and $c_1 = 52$. We take any $\alpha > 0$ and measure $c_{\frac{1}{2}}$ such that $\alpha_{(0, c_{\frac{1}{2}}]}0 \sim \alpha_{(c_{\frac{1}{2}}, 52]}0$. By Observation 1, $C((0, c_{\frac{1}{2}}]) = C((c_{\frac{1}{2}}, 52]) = \frac{1}{2}$. Once we know $c_{\frac{1}{2}}$, we can measure $c_{\frac{1}{4}}$ and $c_{\frac{3}{4}}$ by eliciting indifferences $\alpha_{(0, c_{\frac{1}{4}}]}0 \sim \alpha_{(c_{\frac{1}{4}}, c_{\frac{1}{2}}]}0$ and $\alpha_{(c_{\frac{1}{2}}, c_{\frac{3}{4}}]}0 \sim \alpha_{(c_{\frac{3}{4}}, 52]}0$. It follows that $C(c_{\frac{1}{4}}) = \frac{1}{4}$ and $C(c_{\frac{3}{4}}) = \frac{3}{4}$. In general, we measure subjective midpoints s of time intervals $(q, t]$ by eliciting indifferences $\alpha_{(q,s]}\beta \sim \alpha_{(s,t]}\beta$ $(\alpha > \beta)$. By doing this repeatedly, we can measure the cumulative function C to any desired degree of precision. We can then derive the discount factors from C.

The DM assumes discounted utility. Its most critical property is *separability*: a preference $(x_1, ..., x_{52}) \succcurlyeq (y_1, ..., y_{52})$ with a common outcome $x_i = y_i = c$ is not affected if this common outcome is replaced by another common outcome $x_i = y_i = c'$. By repeated application, preference then is independent of any number of common outcomes.

The next proposition shows that the DM permits a simple test of separability, which we implemented in our experiment. The proposition holds for any outcome $\alpha > 0$ and, more generally, for any pair of outcomes $\alpha > \beta$ with β instead of 0. The proof is in the appendix.

PROPOSITION 3. Under weak ordering and separability, we must have:
(i) $\alpha_{(c_{\frac{1}{4}},c_{\frac{1}{2}}]}0 \sim \alpha_{(c_{\frac{1}{2}},c_{\frac{3}{4}}]}0$;
(ii) $\alpha_{(0,c_{\frac{1}{4}}]}0 \sim \alpha_{(c_{\frac{3}{4}},52]}0$.

2.3 The traditional utility-based method (UM)

Our experiment compares the DM with a traditional *utility(-based) method (UM)*, replicating the implementation by EFB. We first measured prospect theory's utility function from elicited certainty equivalents of 20 risky options. Next, we measured the money amount λ such that

$$90_3 0 \sim \lambda_j 0, \tag{2.5}$$

where $\lambda_j 0$ stands for receiving λ at time (week) j and 0 at all other times. Unlike the DM, the UM only involves one-time payments. We chose $90_3 0$ (and avoided time 0) to have stimuli similar to those of the DM. Using the measured utility function U and Eq.2.1 (discounted utility), we derive from Eq.2.5:

$$\frac{d_j^u}{d_3^u} = \frac{U(90)}{U(\lambda)} \tag{2.6}$$

Here d_j^u is the discrete *utility based discount factor* of week j. We usually normalized $d_3^u = 1$.

2.4 Experiment

2.4.1 Subjects:

We recruited 104 students (61% male; median age 21), mostly economics or finance bachelors, from Erasmus University Rotterdam. The experiment was run at the Econ-Lab of Erasmus School of Economics. The data were collected in five sessions. Seven subjects gave erratic answers[2] and their data were excluded from the analyses.

2.4.2 Incentives:

Each subject was paid a €5 participation fee immediately after the experiment. In addition, we randomly selected (by bingo machine) one subject in each session and then one of his choices to be played out for real. The selections were made in public. We transferred the amount won to the subject's bank account at the dates specified in the outcome streams. In the DM, subjects made choices between streams of money. Consequently, if one of the DM questions was played out for real, we made bank transfers during several weeks. The five subjects who played for real earned €290 on average. Over the whole group, the average payment per subject was €18.70.

2.4.3 procedure:

The experiment was computerized. Subjects sat in cubicles to avoid interactions. They could ask questions at any time during the experiment. The experiment took 45 minutes on average. The first part of the experiment consisted of the DM questions, and the second and third part consisted of the UM questions. Subjects could only start each part after they had correctly answered two comprehension questions. Training questions familiarized subjects with the stimuli.

[2]Debriefings revealed that at least two of these subjects ignored all future payoffs because they had no bank account.

2.4.4 Stimuli: part 1

Part 1 consisted of five questions to measure discounting using the DM and two questions to test separability. To measure discounting, we elicited $c_{\frac{1}{2}}$, $c_{\frac{1}{4}}$, $c_{\frac{3}{4}}$, $c_{\frac{1}{8}}$, and $c_{\frac{7}{8}}$ from the following indifferences:

$$\begin{aligned} &\alpha_{(0,c_{\frac{1}{2}}]}0 \sim \alpha_{(c_{\frac{1}{2}},52]}0, \alpha_{(0,c_{\frac{1}{4}}]}0 \sim \alpha_{(c_{\frac{1}{4}},c_{\frac{1}{2}}]}0, \\ &\alpha_{(c_{\frac{1}{2}},c_{\frac{3}{4}}]}0 \sim \alpha_{(c_{\frac{3}{4}},52]}0, \alpha_{(0,c_{\frac{1}{8}}]}0 \sim \alpha_{(c_{\frac{1}{8}},c_{\frac{1}{4}}]}0, \text{ and } \alpha_{(c_{\frac{3}{4}},c_{\frac{7}{8}}]}0 \sim \alpha_{(c_{\frac{7}{8}},52]}0. \end{aligned} \tag{2.7}$$

To test separability, we measured the indifferences in Proposition 3.

Each question was presented as a choice list in which subjects chose between two options, A and B, in each row. Figure 2.1 displays a screen that subjects faced. In the first choice (first row), B dominates A. Moving down the list, A becomes more attractive and in the final choice A dominates B. The computer enforced monotonicity: After a choice A [B], the computer automatically selected A [B] for all rows below [above], A [B] being more attractive there. Thus, there was a unique switch from B to A between two values. We took the indifference value as the midpoint between these two values. In Figure 2.1, which measures $c_{\frac{1}{8}}$ for a subject who had $c_{\frac{1}{4}} = 13$, the subject switched between 5 and 6 weeks and the indifference value was therefore 5.5.

Figure 2.1: Choice list for the DM elicitation

Which option do you prefer?

Gain €20 per week

Option A	A	B	Option B
in week **1** [1]	○	◉	starting week **1** and ending (after) week **13** [13]
in week **1** [1]	○	◉	starting week **2** and ending (after) week **13** [12]
starting week **1** and ending (after) week **2** [2]	○	◉	starting week **3** and ending (after) week **13** [11]
starting week **1** and ending (after) week **3** [3]	○	◉	starting week **4** and ending (after) week **13** [10]
starting week **1** and ending (after) week **4** [4]	○	◉	starting week **5** and ending (after) week **13** [9]
starting week **1** and ending (after) week **5** [5]	○	◉	starting week **6** and ending (after) week **13** [8]
starting week **1** and ending (after) week **6** [6]	◉	○	starting week **7** and ending (after) week **13** [7]
starting week **1** and ending (after) week **7** [7]	◉	○	starting week **8** and ending (after) week **13** [6]
starting week **1** and ending (after) week **8** [8]	◉	○	starting week **9** and ending (after) week **13** [5]
starting week **1** and ending (after) week **9** [9]	◉	○	starting week **10** and ending (after) week **13** [4]
starting week **1** and ending (after) week **10** [10]	◉	○	starting week **11** and ending (after) week **13** [3]
starting week **1** and ending (after) week **11** [11]	◉	○	starting week **12** and ending (after) week **13** [2]
starting week **1** and ending (after) week **12** [12]	◉	○	in week **13** [1]
starting week **1** and ending (after) week **13** [13]	◉	○	in week **13** [1]

submit my choices

We only used integer-week periods as stimuli to keep the choices simple. Hence, we could not always use the indifference values in subsequent questions and we had to make rounding assumptions. We rounded values below 26 weeks upwards (e.g., 5.5 to 6 weeks), and values above 26 weeks downwards (e.g., 35.5 to 35 weeks) in subsequent choices. Details of our rounding and analyses are in the appendix and online appendix. Our conclusions remained the same under different rounding assumptions, with one exception mentioned later. After each choice list, we asked a control question (explained in Online Appendix WA3).

2.4.5 Stimuli: part 2

Part 2 consisted of seven questions of the type $A = 90_3 0 \sim \lambda_j 0 = B$ with weeks j = 4, 12, 20, 28, 36, 44, and 52. Following EFB, we kept the early outcome in Option A constant and varied the gain λ in Option B (Figure 2.2). As in part 1, the computer enforced monotonicity. EFB only used the timepoints 1 day, 2 months + 1 day, and 4 months + 1 day. We changed these to obtain more detailed measurements and to facilitate comparison with our DM measurements.

Figure 2.2: Choice list for the UM elicitation

Which option do you prefer?

Option A	A	B	Option B
Gain €90 after 3 weeks	○	○	Gain €220 after 36 weeks
	○	○	Gain €210 after 36 weeks
	○	○	Gain €200 after 36 weeks
	○	○	Gain €190 after 36 weeks
	○	○	Gain €180 after 36 weeks
	○	○	Gain €170 after 36 weeks
	○	○	Gain €160 after 36 weeks
	○	○	Gain €150 after 36 weeks
	○	○	Gain €140 after 36 weeks
	○	○	Gain €130 after 36 weeks
	○	○	Gain €120 after 36 weeks
	○	○	Gain €110 after 36 weeks
	○	○	Gain €100 after 36 weeks
	○	○	Gain €90 after 36 weeks
	○	○	Gain €80 after 36 weeks

2.4.6 Stimuli: part 3

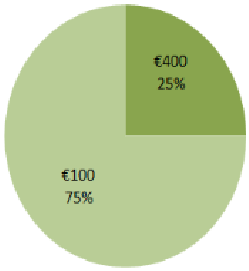
We elicited the certainty equivalents (CE) of twenty risky prospects, shown in Table 2.1, to measure prospect theory's utility function. The CE choice lists appeared in random order. They consisted of choices between sure amounts (option B) and risky prospects (option A) yielding x_1 with probability p and $x_2 < x_1$ otherwise. We used a choice list in which the sure amount that B offered decreased from x_1 in the first row to x_2 in the final row. We used the prospects in EFB with all amounts multiplied by 10 (to get amounts similar to those in the DM), and we used Euros instead of Swiss Francs. Figure 2.3 gives an example of one of the choice lists.

Table 2.1: Risky prospects

p	x_1	x_2	**p**	x_1	x_2
0.10	200	100	**0.25**	500	200
0.50	200	100	**0.50**	500	200
0.90	200	100	**0.75**	500	200
0.05	400	100	**0.95**	500	200
0.25	400	100	**0.05**	1500	500
0.50	400	100	**0.50**	100	0
0.75	400	100	**0.50**	200	0
0.95	400	100	**0.05**	400	0
0.05	500	200	**0.95**	500	0
0.10	1500	0	**0.25**	400	0

Figure 2.3: Choice list of prospects for the CE elicitation

Option A	A	B	Option B
Gain €400 with 25% chance and €100 with 75% chance	○	○	Gain €400 for sure
	○	○	Gain €380 for sure
	○	○	Gain €360 for sure
	○	○	Gain €340 for sure
	○	○	Gain €320 for sure
	○	○	Gain €300 for sure
	○	○	Gain €280 for sure
	○	○	Gain €260 for sure
	○	○	Gain €240 for sure
	○	○	Gain €220 for sure
	○	○	Gain €200 for sure
	○	○	Gain €180 for sure
	○	○	Gain €160 for sure
	○	○	Gain €140 for sure
	○	○	Gain €120 for sure
	○	○	Gain €100 for sure



2.5 Results

Because normality of distributions was always rejected, we used Wilcoxon signed rank tests throughout.

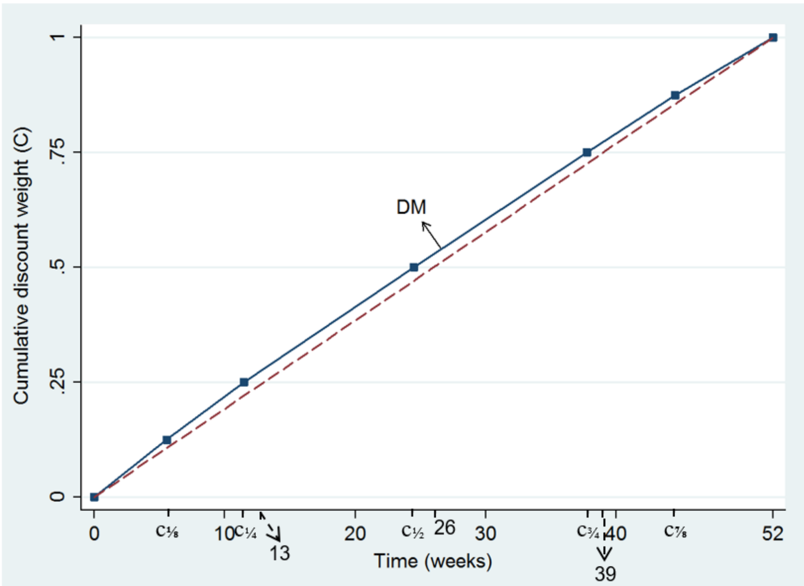
2.5.1 Results for the DM

In all tests reported below $p \leq 0.001$ except when noted. The DM elicits subjective midpoints of time intervals $(q, t]$, denoted $s(q, t]$. Table 2.2 shows that $s(q, t]$ was always closer to q than to t, which is consistent with impatience. Figure 2.4 shows that the cumulative C function was concave, also indicating impatience. We can derive the discount factors $d_j = C(j) - C(j-1)$ from C. They are in Figure 2.5 in Section 2.5.3, where they are compared with the UM discount factors.

Table 2.2: Descriptive statistics of the Direct Method (DM)

Variable	Mean	SD	Min	Median	Max	N
$c_{\frac{1}{8}}$	5.55	1.25	2.13	6.13	9.13	97
$c_{\frac{1}{4}}$	11.47	1.91	4.25	12.25	15.25	97
$c_{\frac{1}{2}}$	24.47	2.72	14.50	25.50	29.50	97
$c_{\frac{3}{4}}$	37.77	2.22	27.75	38.75	42.75	97
$c_{\frac{7}{8}}$	44.5	1.45	39.88	44.88	47.88	97

Figure 2.4: C function of mean data



Statistical tests confirmed the above observations. In all tests, we could reject the one-sided null of no or negative impatience ($s(q,t] \geq \frac{q+t}{2}$) in favor of the alternative hypothesis of impatience ($s(q,t] < \frac{q+t}{2}$) [3].

Decreasing impatience, found in many studies, implies that $\frac{c_{1/2}}{2} - c_{1/4} > \frac{(c_{1/2}+52)}{2} - c_{3/4}$ and $\frac{c_{1/4}}{2} - c_{1/8} > \frac{(c_{3/4}+52)}{2} - c_{7/8}$. The evidence on decreasing impatience was mixed and depended on the rounding assumption used (see the online appendix). Under one rounding assumption[4], we found decreasing impatience in the comparison between $(0, c_{1/2}]$ and $(c_{1/2}, 52]$ and increasing impatience in the comparison between $(0, c_{1/4}]$

[3] $c_{1/2} < 26, c_{1/4} < c_{1/2}/2$, $c_{1/2} < (c_{1/4} + c_{3/4})/2$ (marginally significant), $c_{3/4} < (c_{1/3} + 52)/2$, $c_{1/8} < c_{1/4}/2$, and $c_{7/8} < (52 + c_{3/4})/2$.

[4] A large middle group (n=37) gave answers as close as possible to constant discounting. The first rounding takes them as slightly impatient. It can also be argued that the null of constant discounting should be accepted for them (our second rounding).

and $(c_{3/4}, 52]$. Under another rounding assumption, the null of constant impatience could not be rejected. For all other tests in this paper, the rounding assumptions were immaterial.

To test separability condition (i) in Proposition 3, we directly measured the subjective midpoint $s_{1/2}$ of $(c_{1/4}, c_{3/4}]$. That is, $\alpha_{(c_{1/4}, s_{1/2}]}0 \sim \alpha_{(s_{1/2}, c_{3/4}]}0$. By condition (i), $s_{1/2}$ should equal $c_{1/2}$. To test separability condition (ii) in Proposition 3, we directly measured the value $s_{3/4}$ such that $\alpha_{(0, c_{3/4}]}0 \sim \alpha_{(s_{3/4}, 52]}0$. This second measurement is of a different nature than the questions asked in the rest of the experiment, because now a lower point of an interval is determined rather than a midpoint. By Condition (ii), $s_{3/4}$ should equal $c_{3/4}$.

Separability was rejected in the first test ($p < 0.01$ two-sided), but not in the second. Even in the first test, we found few violations of separability at the individual level. Separability was satisfied exactly for 54 out of 97 subjects. Moreover, $s_{1/2}$ and $c_{1/2}$ differed by 1 at most for 80 subjects. In the second test, separability could not hold exactly due to rounding, but 55 subjects had the minimal difference of 0.5, and 76 subjects had the difference 1.5 or less.

2.5.2 Results for the UM

Table 2.3 summarizes the descriptive statistics of the discount factors d_j^u under the normalization $d_3^u = 1$, the discount factor of the shortest delay in the UM. Comparing pairs of consecutive discount factors confirmed impatience (always $p < 0.001$). We derived the *cumulative function* $C^u(j) = \sum_{i=1}^{j} d_i^u$ from the discount factors. For easy comparison with the DM, we renormalized $C^u(52) = 1$ for C^u.

Table 2.3: Descriptive statistics of the utility-based method (UM)

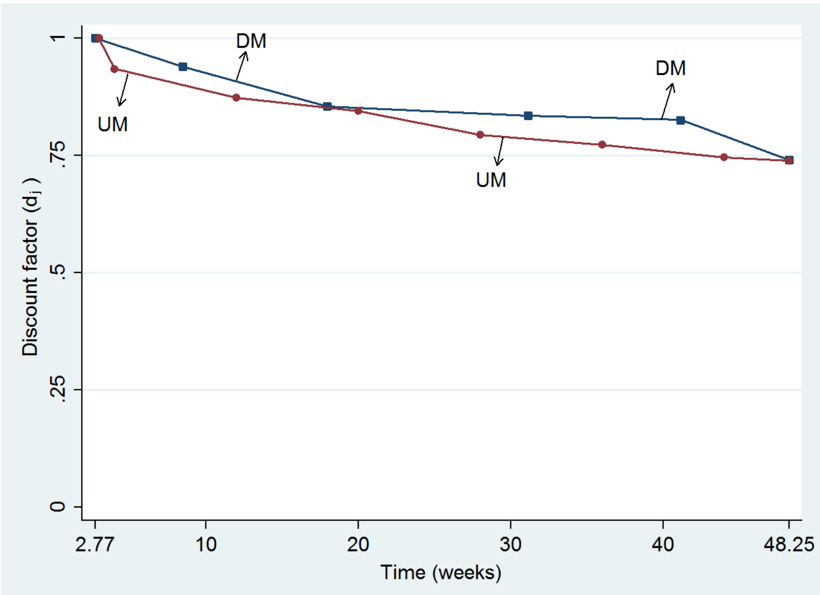
Variable	Mean	SD	Min	Median	Max	N[5]
d_4^u	0.93	0.07	0.67	0.96	1	96
d_{12}^u	0.87	0.12	0.44	0.92	1	96
d_{20}^u	0.85	0.14	0.33	0.89	1	96
d_{28}^u	0.79	0.17	0.33	0.83	1	96
d_{36}^u	0.77	0.18	0.33	0.81	1	96
d_{44}^u	0.75	0.20	0.26	0.78	1	96
d_{52}^u	0.73	0.21	0.26	0.76	1	96

2.5.3 Comparing discounting and impatience under the DM and the UM

Figure 2.5 shows the discount factors of the DM and the UM. For easy comparison, we normalized both to 1 at week 3 here. Both discount factors were decreasing, confirming impatience. The DM discount factors slightly exceeded the UM discount factors, but not significantly (tests provided later). According to both methods, the annual discount rate was 35%, assuming continuous compounding e^{-rt} with t in years.

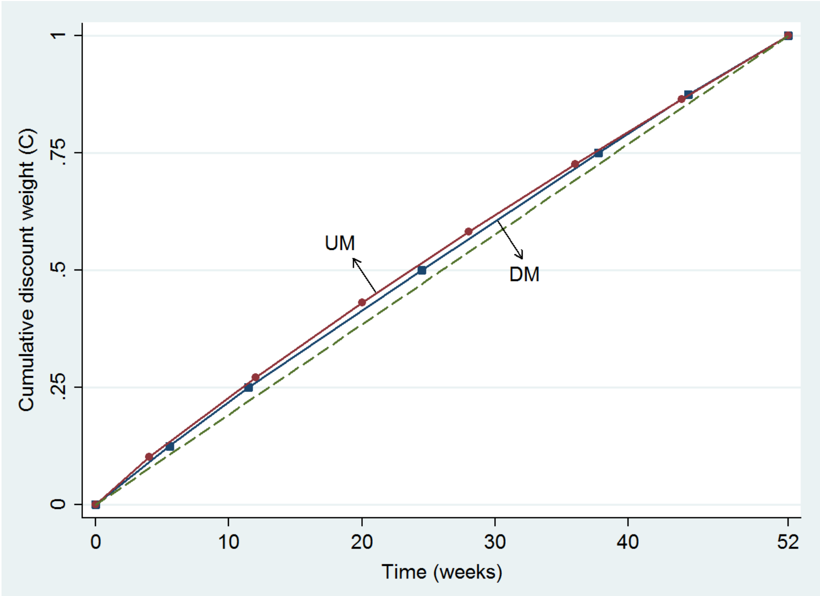
Figure 2.6 shows the cumulative functions of the DM and the UM, using linear interpolation to obtain general values d_j^u from the seven measured d_j^u's. We measured impatience (= concavity) for each cumulative function by the difference between the area under this function and that under the diagonal ($t \mapsto t/52$). This index is negative for convex functions. It is 0 if the decision maker does not discount.

Figure 2.5: Comparing discount factors of the DM and the UM



[5] We excluded one subject because of his extreme power (-118.7; overall average is 0.53) for utility.

Figure 2.6: Comparing cumulative discount weights of the DM and the UM



The average value of the impatience index was 1.14 for the DM and 1.62 for the UM. Both indices exceeded 0 ($p < 0.001$), in agreement with impatience. The index for the UM exceeded that for the DM ($p = 0.02$), suggesting more impatience under the UM.

We also estimated the DM and UM curves in Figure 2.6 by a power function. The median powers were 0.96 for the DM and 0.93 for the UM. Both powers were below 1 (both $p < 0.001$), again confirming impatience, but they did not differ significantly from each other. The three measures of impatience (discount factors for week 52, area-differences, and estimated power coefficients) correlated strongly (≥ 0.90), for both the DM and the UM. For consistency between the two methods at the individual level, we tested correlations of the three measures of impatience between the DM and UM. They were all around 0.25 ($p < 0.001$). Hence, even though the different measures led to consistent conclusions within methods, differences remained.

2.5.4 Parametric estimations

The discount factors trace out the discount function D(t) without making parametric assumptions (see Figure 2.5). This section reports parametric fittings. We estimated the discount function of each subject by maximum likelihood using the following three parametric families.

1. *Constant discounting* (Samuelson 1937), with one parameter $r \geq 0$:

$$D(t) = e^{-rt} \text{ with } r \geq 0.$$

2. *Hyperbolic discounting* (Loewenstein and Prelec 1992), with two parameters $\alpha \geq 0$ and $\beta > 0$:

$$\text{For } \alpha > 0 : D(t) = (1 + \alpha t)^{-\frac{\beta}{\alpha}};$$

$$\text{for } \alpha = 0 : D(t) = e^{-\beta t}.$$

   The α parameter determines how much the discount function departs from constant discounting. The limiting case $\alpha = 0$ reflects constant discounting. The β parameter determines impatience.

3. *Unit-invariance discounting*[6] , with two parameters $r > 0$ and d:

$$\text{For } d > 1, D(t) = e^{rt^{1-d}} \text{ (only if t = 0 is not considered);}$$

$$\text{for } d = 1, D(t) = t^{-r} \text{ (only if t = 0 is not considered);}$$

$$\text{for } d < 1, D(t) = e^{-rt^{1-d}}.$$

   The r-parameter determines impatience and the d-parameter determines departure from constant discounting, interpreted as sensitivity to time by Ebert and Prelec (2007). The common empirical finding is $d \leq 1$, reflecting insensitivity.

Hyperbolic discounting can only account for decreasing impatience. However, empirical studies have observed that a substantial proportion of subjects are not decreasingly, but increasingly impatient (references in section 2.5.5). Unit-invariance discounting can account for both decreasing and increasing impatience. We can use the entire

[6] Read (2001 Eq. 16) first suggested this family. Ebert and Prelec (2007) called it constant sensitivity, and Bleichrodt et al. (2009) called it constant relative decreasing impatience. Bleichrodt et al. (2013) proposed the term unit-invariance.

unit invariance family because our domain does not contain $t = 0$ (explained in the discussion section). The exclusion of $t = 0$ also implies that the popular quasi-hyperbolic family coincides with constant discounting for our stimuli.

Table 2.4 shows the estimated parameters. The exponential discounting parameters differed between the UM and the DM ($p < 0.001$), reflecting more discounting for the UM. The parameters of hyperbolic discounting and unit-invariance discounting did not differ significantly ($p > 0.2$).

Table 2.4: Results of parametric fittings for the DM and the UM

		Exponential		**Hyperbolic**		**Unit-Invariance**	
		Mean(r)	Median(r)	Mean(α, β)	Median(α, β)	Mean(d,r)	Median(d,r)
Par.	**UM**	0.009	0.006	1.88, 0.25	1.21, 0.06	0.58, 2.19	0.94, 0.69
	DM	0.005	0.002	1.66, 0.14	1.30, 0.05	0.80, 1.86	0.89, 0.25
AIC	**UM**	-3.03	-3.26	-1.55	-1.68	-1.63	-1.70
	DM	-3.56	-3.68	-1.66	-1.80	-1.65	-1.78

The final two rows of Table 2.4 show the goodness of fit of the three discount families by the Akaike information criterion (AIC). More-negative values indicate better fit. The DM method with exponential discounting fitted better than the UM with exponential discounting ($p < 0.001$), and gave the best fit overall. The DM also seemed to fit better for hyperbolic discounting and unit-invariance, but these differences were not significant. Of the three parametric families, exponential discounting fitted best for both the DM and the UM (both $p < 0.001$). For the UM, hyperbolic discounting gave the worst fit ($p < 0.001$). For the DM we found no significant difference between unit-invariance and hyperbolic discounting ($p = 0.22$). In the absence of the immediacy effect, exponential discounting performed well, which also supports quasi-hyperbolic discounting. The online appendix gives further details.

We, finally, investigated the relation between impatience (concavity of C and C^u) and risk attitudes, controlling for demographic variables (age, gender, and foreign versus domestic-Dutch). Impatience under the UM was negatively related with concavity of utility, which is not surprising because the UM measurements were based on utility. Under the DM, impatience was not related with utility, suggesting that these are independent components. Impatience under the UM was also negatively related with risk aversion in the form of pessimism of probability weighting, in contrast to impatience under the DM. Age was positively related with UM impatience. Other relations were not significant. Details are in the online appendix.

2.5.5 Discussion of the results

The DM and the UM led to similar conclusions. Under both methods, subjects were impatient. However, we found less discounting with the DM. The high estimated annual discount rate (35%) suggests that this is a desirable feature of the DM.

Even though theoretical studies commonly assume universal decreasing impatience, many empirical studies have found considerable increasing impatience at the individual level. We found prevailing decreasing impatience in the UM, but mixed evidence in the DM. Statistical tests only showed weak evidence for decreasing impatience, and Figure 2.5 suggests that impatience was not always decreasing. Increasing impatience implies that people become more reluctant to wait as time passes by. The presence of several subjects with increasing impatience in our data also explains the poor performance of the hyperbolic discount functions, which only allow for universal decreasing impatience, and cannot fit the data of increasingly impatient subjects.

Our measurement of the DM included two tests of separability, a condition underlying the method. One test suggested violations of separability, but we could not reject separability in the other test. In both tests, most subjects behaved in agreement with it. The DM, like any decision model, does not fit data perfectly, but we still use such decision models in the absence of better models that are sufficiently tractable. Violations of separability may, for example, be due to sequencing effects and habit formation (Loewenstein and Prelec, p.350; Dolan and Kahneman, p.228). The DM allows easy tests of separability that help to assess its restrictiveness. Such tests are desirable because separability is used in virtually all applications of discount measurements.

Besides separability, our analysis also assumes independence of discounting from the outcome used. This condition is sometimes called separability of money and time, and its violation the magnitude effect (Loewenstein and Prelec 1992). If magnitude effects exist, then our measurements are only valid for outcomes close to those used in the measurements.

We used the method of EFB for the UM. Our estimates of risk attitudes were close to theirs except for the curvature of utility, which can be explained by the larger outcomes we used (further details are in the appendix). We could not directly compare our findings on discounting with those of EFB, because they used fewer and different timepoints. The negative relation between concave utility and impatience that we found for the UM is not surprising because utility plays a central role in the UM. Concave utility increases the ratio in Eq. 2.6 and thus decreases impatience. The negative relation between impatience and probability weighting suggests that this component of risk attitude also affects the UM measurements. Our findings suggest that there is collinearity between utility/risk attitude and discounting in the UM, but not in the DM.

Our implementation of the DM is adaptive, with answers to questions influencing the stimuli in later questions. Theoretically, this may offer scope for manipulation: responding untruthfully to some questions may improve later stimuli. However, according to Bardsley et al. (2010 p. 265) classification, this possibility is only theoretical and is no cause for concern in our experiment. First, it was virtually impossible for subjects to realize that questions were adaptive because of the roundings used. Second, we expect that even readers like us, who are aware of the adaptive nature and even of the stimuli of our experiment beforehand, are not likely to see how the loss of wrongly answering one question could be compensated by advantages in follow-up questions. This would be impossible for our subjects. Online Appendix WE gives details. For these two reasons, manipulation is only a theoretical concern for our experiment.

The DM can be implemented nonadaptively. For example, we can select a number of outcome streams $\alpha_{A^j}\beta$ and timepoints s^j beforehand, and measure the timepoints t^j such that $\alpha_{A^j}\beta \sim \alpha_{(s^j,t^j]}\beta$ for all j. Observation 1 still gives equalities $C(A^j) = C(s^j, t^j]$. We can use these in parametric fittings of C or in tests of properties such as decreasing impatience, without requiring knowledge about U. A drawback of this nonadaptive procedure is that we then cannot readily draw a connected C-curve as in Figures 2.4 and 2.6, where we needed no parametric assumption (other than linear interpolation).

The DM always had a better fit than the UM, and exponential discounting always fitted best, with unit-invariance second best. Exponential discounting could perform well because we did not include the present $t = 0$ in our stimuli, where most violations are found due to the immediacy effect (Attema, 2012). Although this effect is important and deserves further study, we decided to focus our first implementation of the DM on a better understood empirical domain, which we could compare directly with EFB. In this regard, we follow many other studies in the literature that use front-end delays.

The DM can readily investigate the immediacy effect and discrete outcomes at $t = 0$. The latter are then interpreted as salaries paid at the beginning, instead of at the end of periods (weeks in our case). Given that the relations that we made with flow variables only served as intermediate tools in our mathematical analysis, and played no role in the stimuli or results, we can use the interpretation mentioned. Quasi-hyperbolic discounting then implies a high weight for the first week of salary (now mathematically representing the present rather than the timepoint one week ahead), and moderate weights for the other weeks.

2.6 General discussion

Attema et al. (2010) measured discounting up to a power without the need to know utility, but needed separate measurements to identify the power. Unlike the DM but like the UM, their measurements did not need time separability, but they could not test it either. In a mathematical sense, our method is similar to the measurement of subjective probability based on equal likelihood assessments, where utility also drops from the equations and separability (now over events) is assumed (Baillon, 2008).

Subjective midpoints, used by the DM to measure discounting, have a long tradition in psychophysics (bisection; Stevens, 1936) and mathematics (quasi-arithmetic mean; Aczel, 1966). Condition (i) in Proposition 3, a necessary condition of a quasi-arithmetic mean, is a special case of autodistributivity (Aczel, 1966; Eq. 6.4.2.3, for t the midpoint of x and y).

To our best knowledge, all experimental measurements of money discounting have used discrete outcomes. Real-life decisions often involve flow outcomes that are repeated per time unit. Examples are salary payments, pension saving plans, and mortgage debt repayments. In such contexts, the DM is more natural than discrete methods such as the UM. For discrete outcomes, the DM can be an alternative to the UM if the payments are sufficiently frequent and the periods are sufficiently fine, as in our experiment. However, the DM is less useful for decisions in which outcomes occur infrequently, such as for single-outcome decisions.

In the DM, subjects only make tradeoffs between periods. In the UM, subjects make tradeoffs both between outcomes and between periods, which is more complex. Hence, the DM is easier for subjects. Our experiment gave indirect support: We found a positive correlation between utility curvature and discounting for the UM, but not for the DM, showing that outcome tradeoffs impact time tradeoffs in the UM, but not in the DM.

The DM is also easier to use for researchers, because of the elementary nature of Observation 1. In the DM, we only used seven questions[7]. In the UM, we also used seven questions to elicit discounting, but we needed additional questions to elicit utility. The DM took much less time.

The DM can be analyzed using parametric econometric fittings (Section 2.5.4), as can all existing methods, but, unlike most methods, the DM can also be analyzed in a parameter-free way (Section 2.5.1). This reveals the correct discount function without a commitment to a parametric family of discount functions. The DM can also be used for interactive prescriptive measurements in consultancy applications (Keeney and Raiffa, 1976).

2.7 Conclusion

This paper has introduced a new method to measure the discounting of money, the direct method. This method is simpler than existing methods because it does not need information about utility. Consequently, the experimental tasks are easier for subjects, researchers have to ask fewer questions, and the measurements are not distorted by biases in utility. An experiment confirms the implementability and validity of the

[7]We used the same numbers of questions to make the methods comparable. In fact, two DM questions tested separability. The DM derived the discount functions from only five measurements.

direct method.

2.8 Appendix A. Proofs

PROOF OF PROPOSITION 3. This proof is elementary in not using technical assumptions such as continuity. The proof only uses the two outcomes used in the preferences, being α and 0. For deriving the first indifference, we denote outcome streams as quadruples (x_1, x_2, x_3, x_4), with x_1 received in $(0, c_{\frac{1}{4}}]$, x_2 in $(c_{\frac{1}{4}}, c_{\frac{1}{2}}]$, x_3 in $(c_{\frac{1}{2}}, c_{\frac{3}{4}}]$, and x_4 in $(c_{\frac{3}{4}}, 52]$. We only use the following parts of Eq. 2.7: $(\alpha, \alpha, 0, 0) \sim (0, 0, \alpha, \alpha)$ (1), $(\alpha, 0, 0, 0) \sim (0, \alpha, 0, 0)$ (2), and $(0, 0, \alpha, 0) \sim (0, 0, 0, \alpha)$ (3). Assume, for contradiction, that (i) is violated, say $(0, \alpha, 0, 0) \succ (0, 0, \alpha, 0)$ (4). This, (2), (3), and transitivity, imply $(\alpha, 0, 0, 0) \succ (0, 0, 0, \alpha)$ (5). By separability, (4) implies $(\alpha, \alpha, 0, 0) \succ (\alpha, 0, \alpha, 0)$, and (5) implies $(\alpha, 0, \alpha, 0) \succ (0, 0, \alpha, \alpha)$. By transitivity, $(\alpha, \alpha, 0, 0) \succ (0, 0, \alpha, \alpha)$, contradicting (1). Reversing all preferences shows that $(0, \alpha, 0, 0) \prec (0, 0, \alpha, 0)$ implies the contradictory $(\alpha, \alpha, 0, 0) \prec (0, 0, \alpha, \alpha)$. The indifference in (i) has been proved. The second indifference follows from the first, (2), (3), and transitivity.

2.9 Appendix B. Details of the DM

Preferences $\alpha_{\{1,\ldots,j\}}0 \prec \alpha_{\{j+1,\ldots,52\}}0$ and $\alpha_{\{1,\ldots,j+1\}}0 \succ \alpha_{\{j+2,\ldots,52\}}0$ reveal that $c_{\frac{1}{2}}$ is in the interval $(j, j+1)$. We then estimate $c_{\frac{1}{2}} = j + \frac{1}{2}$. For the DM, we used the following roundings to derive the discount factors from the C function (Figure 2.5). For each of the six periods considered (bounded by $t = 0$, the five c_p values that we measured, and $t = 52$), we divided the increase of C over this period by the length of the period to obtain the average week-weight d over this period. We assigned this d value to the midpoint of the period, and we used linear interpolation between these midpoints. We normalized (setting $d = 1$) at the smallest positive timepoint considered, being $\frac{c_{1/8}}{}2$. Its average (2.75) was approximately 3, leading to about the same normalization as with the UM. Thus we obtained a d-function over the interval (3, 48.25], with 48.25 the average midpoint of the last interval $(c_{\frac{7}{8}}, 52]$.

Because we only presented integer-week periods to subjects, and estimates of c_p were usually nonintegers, we could not present exact c_p values to our subjects in our adaptive experiment. For example, to find the subjective midpoint $c_{\frac{1}{4}}$ of $(0, c_{\frac{1}{2}}]$, we rounded $c_{\frac{1}{2}}$ and took the smallest larger integer, denoted $j+1$ here, and then found the subjective midpoint x of $(0, j+1]$. To derive $c_{\frac{1}{4}}$ from this midpoint x, we corrected for the roundings. Because we had used $j+1$ instead of $c_{\frac{1}{2}}$, which on average is an overestimation of $c_{\frac{1}{2}}$ by $\frac{1}{2}$, and half of it will propagate into x, we subtracted $\frac{1}{4}$ from x to get $c_{\frac{1}{4}}$. In all other estimations of values c_p, we similarly used roundings and corrections. Complete details of the roundings and corrections for all c_p are in the online appendix.

2.10 Appendix C. Details of the UM

Following EFB, we adopted power utility Wakker (2008) and Prelec's (1998) two-parameter probability weighting:

$$\text{If } \eta > 0\text{, then } u(x) = x^{\eta};$$

$$\text{If } \eta = 0\text{, then } u(x) = ln(x);$$

$$\text{If } \eta < 0\text{, then } u(x) = -x^{\eta};$$

$$w(p) = e^{-\beta(-ln(p))^{\alpha}}.$$

For convenience, x is generic for outcomes in this appendix. The average value of the utility parameter η was 0.47 (in EFB, $\eta = 0.87$), which reflects concavity[8]. The difference found can be explained by the higher outcomes we used. The average insensitivity index α was 0.55 (in EFB, $\alpha = 0.51$), indicating departure from linear probability weighting. The average estimates for the pessimism index β was 0.94 (in EFB, $\beta = 0.97$). In Eq. 2.6 we should have $U(0) = 0$, which agrees with prospect theory's scaling. Following EFB, the estimation of utility is carried out after shifting all outcomes by one unit of money, so as to avoid mathematical complications of logarithmic or negative-power utility at $x = 0$. We followed EFB in using choice lists to elicit λ in Eq. 2.5 (details are in the online appendix). If the largest value in a choice list was still too small to lead to preference, we assumed preference to switch in the

[8]The average value of the parameters in our analysis is based on 96 subjects (including subjects who have missing values). EFB removed all subjects with missing values.

first higher value to follow. We thus use censored data. It gives a smaller bias than dropping these subjects, the most impatient ones, as done by EFB, and it keeps more subjects for other measurements. The DM measurements need no censoring of data because the indifference points are always between extremes of the choice lists.

Chapter 3

An Experimental Test of Reduction Invariance

3.1 Introduction

Probability weighting is an important reason why people deviate from expected utility (Fox and Poldrack, 2014; Luce, 2000; Wakker, 2010). Prelec (1998) proposed a functional form for the probability weighting function that is widely used in empirical research and usually gives a good fit to empirical data (Sneddon and Luce, 2001; Stott, 2006; Chechile and Barch, 2013).

Although other functional forms have also been used (e.g. Currim and Sarin, 1989; Goldstein and Einhorn, 1987; Karmarkar, 1978; Lattimore et al., 1992; Tversky and Kahneman, 1992), Prelec was the first to give an axiomatic foundation for a form of the probability weighting function[1]. His central condition, compound invariance (defined in Section 3.2), is, however, complex to test empirically as it involves four indifferences and may be subject to error cumulation. To the best of our knowledge, it has not been tested yet.

Luce (2001) proposed a simpler condition, reduction invariance. Luce (2000, p.278)

This chapter is based on the homonymous paper, co-authored with Ilke Aydogan and Han Bleichrodt.

[1]For a more recent axiomatic analysis of probability weighting see Diecidue et al. (2009).

identified testing reduction invariance as an important open empirical problem. The purpose of this paper is to follow up on Luce's suggestion and to test reduction invariance in an experiment. Our data support the validity of reduction invariance. At the aggregate level, we found evidence for the condition and at the individual level it was clearly the dominant pattern.

A special case of reduction invariance is the rational case of reduction of compound gambles, which implies that the probability weighting function is a power function. Our data on reduction of compound gambles are mixed. At the aggregate level reduction of compound gambles was clearly violated. However, 60% of our subjects behaved in line with it. The subjects who deviated, did so systematically and found compound gambles more attractive than simple gambles.

3.2 Background

Let (x, p) denote a *gamble* which gives consequence x with probability p and nothing otherwise. Consequences can be pure, such as a money amounts, or they can be a gamble (y, q) where y is a pure consequence. The set of pure consequences is a nonpoint interval $\mathcal{X}$ in $\mathbb{R}^+$ that contains 0. Preferences $\succcurlyeq$ are defined over the set $\mathcal{C}$ of gambles. We identify preferences over simple gambles (x, p) from preferences over $((x, p), 1)$ and preferences over consequences x from preferences over $(x, 1)$.

A function U *represents* $\succcurlyeq$ if it maps gambles and pure consequences to the reals and for all gambles $(x, p), (x', p')$ in $\mathcal{C}$, $(x, p) \succcurlyeq (x', p') \Leftrightarrow U(x, p) \geq U(x', p')$. If a representing function U exists then $\succcurlyeq$ must be a weak order: transitive and complete. The representing function U is *multiplicative* if there exists a functions $W : [0, 1] \to [0, 1]$ such that:

i. $U(x, p) = U(x)W(p)$.

ii. $U(0) = 0$ and U is continuous and strictly increasing.

iii. $W(0) = 0$ and W is continuous and strictly increasing.

The functions U and W are unique up to different positive factors and a joint positive power: $U \to a_1 U^b$ and $W \to a_2 W^b, a_1, a_2, b > 0$. This uniqueness implies that we can always normalize W such that $W(1) = 1$[2]. Luce (1996, 2000; Marley and Luce,2002) gave preference foundations for the multiplicative representation. A central condition in these results is consequence monotonicity, which we also assume here[3].

The multiplicative representation is general and contains many models of decision under risk as special cases. Examples are expected utility, rank-dependent utility (Quiggin, 1981, 1982), prospect theory (Tversky and Kahneman, 1992), disappointment aversion theory (Gul, 1991), and rank-dependent utility (Luce, 1991; Luce and Fishburn, 1991, 1995).

Prelec (1998) axiomatized the following family of weighting functions:

Definition 1: $W(p)$ is compound-invariant if there exist $\alpha > 0$ and $\beta > 0$ such that $W(p) = exp(-\beta(-lnp)^{\alpha})$.

Prelec's compound-invariant weighting function has several desirable properties. First, it includes the power functions $W(p) = p^{\beta}$ as a special case. The class of power weighting functions is the only one that satisfies reduction of compound gambles, which is often considered a feature of rational choice:

$$((x, p), q) \sim (x, pq).$$

A second advantage of the compound-invariant family is that for $\alpha < 1$, it can account for inverse S-shaped probability weighting, which has commonly been observed in empirical studies (Fox and Poldrack, 2014; Wakker, 2010). Finally, the parameters α and β have an intuitive interpretation (Gonzalez and Wu, 1999). The parameter α reflects a decision maker's sensitivity to changes in probability, with higher values representing more sensitivity, while β reflects the degree to which a decision maker is averse to risk, with higher values reflecting more aversion to risk.

[2]Aczel and Luce (2007) analyzed the case where $W(1) \neq 1$ to model non-veridical responses in psychophysical theories of intensity (Luce, 2002, 2004).

[3]Consequence monotonicity means that if two gambles differ only in one consequence, the one having the better consequence is preferred. As Luce (2000, p. 45) points out, it implies a form of separability for compound gambles. It also implies backward induction, where each simple gamble in a compound gamble can be replaced by its certainty equivalent. von Winterfeldt et al. (1997) found few violations of consequence monotonicity for choice-based elicitation procedures, as used in our experiment, and what there was seemed attributable to the variability in certainty equivalence estimates.

The compound-invariant family of weighting functions satisfies the following condition:

Definition 2: Let N be any natural number. *N-compound invariance* holds if $(x, p) \sim (y, q)$, $(x, r) \sim (y, s)$, and $(x', p^N) \sim (y', q^N)$ imply $(x', r^N) \sim (y', s^N)$ for all nonzero consequences x, y, x', y' and nonzero probabilities p, q, and r.

Compound invariance holds if N-compound invariance holds for all N. Prelec (1998) showed that if compound invariance is imposed on top of the multiplicative representation then $W(p)$ is compound-invariant. Bleichrodt et al. (2013) showed that compound invariance by itself implies the multiplicative representation and, consequently, that the assumption of a multiplicative representation is redundant.

Compound invariance is difficult to test empirically. It requires four indifferences and elicited values appear in later elicitations, which may lead to error cumulation. For example, we could fix x, p, q, r, and x'. The first indifference would then elicit y, the second s, and the third y'. If each of these variables is measured with some error then this will affect the final preference between (x', r^N) and (y', s^N).

To address the problem of error cumulation, Luce (2001) proposed a simpler condition.

Definition 3: Let N be any natural number. *N-reduction invariance* holds if $((x, p), q) \sim (x, r)$ implies $((x, p^N), q^N) \sim (x, r^N)$ for all nonzero consequences x and nonzero probabilities p, q, and r.

Reduction invariance holds if N-reduction invariance holds for all N. Reduction invariance is easier to test than compound invariance as it requires only two indifferences. Luce (2001, Proposition 1) showed that if N-reduction invariance for $N = 2, 3$ is imposed on top of the multiplicative representation then the weighting function $W(p)$ is compound-invariant[4].

[4]To the best of our knowledge, Bleichrodt et al.'s (2013) result cannot be generalized to reduction invariance and the multiplicative representation still has to be assumed in this case.

3.3 Experiment

The purpose of our experiment was to test reduction invariance (for N=2,3) to obtain insight into the descriptive validity of the compound-invariant weighting function. The simplest way to test reduction invariance would be to fix x, p, and q, to elicit the probability r such that a subject is indifferent between $((x,p),q)$ and (x,r), and then to check whether he is indifferent between $((x,p^N),q^N)$ and (x,r^N). However, as Luce (2001) pointed out, a danger of this procedure is that many subjects may realize that $r = pq$ is a sensible response. This may distort the results as empirical evidence suggests that subjects do not satisfy reduction of compound gambles (Abdellaoui et al., 2015; Bar-Hillel, 1973; Bernasconi and Loomes, 1992; Keller, 1985; Slovic, 1969). Luce (2001) suggested another approach for testing reduction invariance, which we adopted in our experiment. Instead of asking for probability equivalents, we elicited the certainty equivalents of $((x,p),q)$, denoted $CE((x,p),q)$, and several $CE(x,r)$ for a range of values of r centered on pq. Using interpolation, we then determined the value r_1 for which $CE((x,p),q) = CE(x,r_1)$. We then elicited $CE((x,p^2),q^2)$ and $CE((x,p^3),q^3)$ and tested whether $CE((x,p^2),q^2) = CE(x,(r_1)^2)$ and $CE((x,p^3),q^3) = CE(x,(r_1)^3)$ where $CE(x,(r_1)^2)$ and $CE(x,(r_1)^3)$ were, again, determined using interpolation.

3.3.1 Procedure

The experiment was run on computers. Subjects were seated in cubicles with a computer screen and a mouse and could not communicate with each other. Once everyone was seated, the instructions were displayed, followed by three comprehension questions. Subjects could only proceed to the actual experiment when they had correctly answered all three comprehension questions. Copies of the instructions and the comprehension questions are in Appendix A.

We measured the certainty equivalents of 12 compound gambles and of 6 simple gambles. The order in which these gambles were presented was random. The winning amount was always €200. Table 3.1 displays the compound gambles that we used. Compound gambles $C1-C4$ were the original gambles, gambles $C5-C8$ were derived

Table 3.1: The compound gambles used in the experiment

Compound gambles	**Gamble**	**Type**	**Reduced probability**	**Expected value**
C1	((€200,82%),67%)	Original	54.94%	€109.88
C2	((€200,45%),67%)	Original	30.15%	€60.30
C3	((€200,63%),90%)	Original	56.70%	€113.40
C4	((€200,82%),39%)	Original	31.98%	€63.96
C5	((€200,67%),45%)	Square of C1	30.15%	€60.30
C6	((€200,20%),45%)	Square of C2	9.00%	€18.00
C7	((€200,40%),81%)	Square of C3	32.40%	€64.80
C8	((€200,67%),15%)	Square of C4	10.05%	€20.10
C9	((€200,55%),30%)	Cube of C1	16.50%	€33.00
C10	((€200,9%),30%)	Cube of C2	2.70%	€5.40
C11	((€200,25%),73%)	Cube of C3	18.25%	€36.50
C12	((€200,55%),6%)	Cube of C4	3.30%	€6.60

from $C1 - C4$ by taking the squares of the probabilities, and gambles $C9 - C12$ were derived from $C1 - C4$ by taking the cubes of the probabilities. Because taking the square and the cube of probabilities usually does not give round numbers, we selected the probabilities in the compound gambles $C1 - C4$ such that only little rounding was necessary in the derived compound gambles. We could have avoided rounding altogether by presenting fractions. However, we observed in the pilot sessions that subjects found complex fractions harder to handle than probabilities.

By comparing the certainty equivalents of $C2$ and $C5$ and (roughly) those of $C4$ and $C7$ we could test whether subjects preferred to have most of the uncertainty resolved in the first stage or in the second stage. Luce (1990, p. 228) already drew attention to modeling the order in which events are carried out and Ronen (1973) and Budescu and Fischer (2001) found that people prefer gambles with high first-stage probabilities and lower second-stage probabilities to gambles with high second-stage probabilities and lower first-stage probabilities. On the other hand, Chung et al. (1994) concluded that with a choice-based procedure most subjects were indifferent to the order in which events were carried out.

Table 3.2 shows the simple gambles that we used in the experiment. The probabilities in the simple gambles were close to the reduced probabilities of the compound gambles.

Table 3.2: The simple gambles used in the experiment

Simple gambles	Gamble	Expected value
S1	(€200,3%)	€6
S2	(€200,9%)	€18
S3	(€200,17%)	€34
S4	(€200,32%)	€64
S5	(€200,57%)	€114
S6	(€200,77%)	€154

To determine the certainty equivalents of the compound and the simple gambles, subjects made a series of choices between these gambles and sure amounts of money. Simple risk and compound risk were represented by urns containing colored balls. The color of the ball determined subjects' payoffs. We used one urn for the simple gambles and two urns for the compound gambles. Appendix A displays the way the simple and the compound gambles were presented.

All certainty equivalents were elicited using a choice-based iterative procedure, which is close to the PEST procedure used by, amongst others, Cho and Luce (1995) and Cho et al. (1994). We did not ask subjects directly for their certainty equivalents as this tends to lead to less reliable measurements (Bostic et al., 1990), but instead used a series of choices to zoom in on them. The iteration procedure is described in Appendix B.

We included two types of consistency tests. First, we repeated the third choice in the iteration procedure for four randomly selected questions. Subjects were usually close to indifference in the third choice and, consequently, this was a rather strong test of consistency. Second, we repeated the entire elicitation of two certainty equivalents, one for a randomly selected simple gamble and one for a randomly selected compound gamble.

3.3.2 Subjects and incentives

The experiment was performed at the ESE-Econlab at Erasmus University in 5 group sessions. Subjects were 79 Erasmus University students from various academic disciplines (average age 23.4 years, 43 female). We paid the subjects a €5 participation fee. In addition, at the conclusion of each session we randomly selected two subjects who could play out one of their randomly drawn choices for real. If a subject had chosen the sure amount in that choice then we paid him that amount. If he had chosen the simple or the compound gamble then we created the relevant urn(s) and the subject drew the ball that determined his payoffs. The 10 subjects who played out one of their choices for real earned on average €49.60 per person. Sessions lasted 45 minutes on average including 10 minutes to implement payment.

3.3.3 Analysis

To test reduction invariance, we followed Luce's (2001) suggestion. We determined for each compound gamble $((€200, p), q)$ the probability r such that $CE((€200, p), q) = CE(€200, r)$ using the certainty equivalents of the simple gambles and linear interpolation. Subjects' certainty equivalents of the simple gambles did not always increase with the probability of winning €200 and, consequently, the value of r for which $CE((€200, p), q) = CE(€200, r)$ could not always be uniquely determined. If there were multiple values of r for which $CE((€200, p), q) = CE(€200, r)$ then we used the average of these values in our analysis. We also analyzed the results using only those responses for which r could be uniquely determined, but this did not affect our conclusions. Finally, we also estimated the weighting function by smoothing splines (Hastie et al.,2008, Section 5.4) and used this estimation to predict r[5]. We discuss the results of this nonparametric regression analysis in the subsection Robustness analysis.

People's preferences are typically stochastic and the elicited certainty equivalents are subject to noise. Moreover, the choice-based procedure determined certainty equivalents up to €1 precision and it was in theory possible that the absolute difference between $CE((€200, p^N), q^N)$ and $CE(€200, r^N), N = 2, 3$, was equal to 2 even though a subject satisfied reduction invariance exactly. For these reasons and because $CE(€200, r^N), N = 2, 3$, had to be approximated, which introduced further imprecision, we considered a test of equality of the certainty equivalents too stringent. Instead, we followed Cho and Luce's (1995) approach in testing preference conditions and compared the pro-

[5]For these estimations we used the smooth.splines function in R (R core team 2015) which estimates prediction error by generalized cross-validation.

portions of respondents for whom $CE((€200, p^N), q^N) > CE(€200, r^N)$ with those for whom $CE((€200, p^N), q^N) < CE(€200, r^N), N = 2, 3$. Under reduction invariance with random error, deviations from equality between $CE((€200, p^N), q^N)$ and $CE(€200, r^N)$ should be nonsystematic and we should observe that the proportion of subjects for whom $CE((€200, p^N), q^N) > CE(€200, r^N)$ does not differ systematically from the proportion for whom $CE((€200, p^N), q^N) < CE(€200, r^N)$. Because our elicitation method only determined certainty equivalents up to €1 precision we took $CE((€200, p^N), q^N)$ and $CE(€200, r^N)$ equal if $|CE((€200, p^N), q^N) - CE(€200, r^N)| \leq 2$[6]. We also analyzed the results using the exact equality. This did not affect our conclusions at the aggregate level but, obviously decreased support for reduction invariance at the individual level[7].

Our null hypothesis is that reduction invariance holds, which involves testing the invariance $P(CE((€200, p^N), q^N) > CE(€200, r^N)) = P(CE((€200, p^N), q^N) < CE(€200, r^N))$. As pointed out by Rouder et al. (2012, 2009) classic null-hypothesis significance tests are less suitable when testing for invariances for two reasons. First, they do not allow researchers to state evidence for the null hypothesis and, second, they overstate the evidence against the null hypothesis. We therefore used Bayes factors to test our null hypotheses. The Bayes factors describe the relative probability of the observed data under the null and the alternative hypothesis. For example, a Bayes factor of 10 will indicate that the null is 10 times more likely than the alternative given the data. We used the package BayesFactor in R (Morey et al., 2015) to compute the Bayes factors. Following Jeffreys (1961) we interpret a Bayes factor larger than 3 as "some evidence" for the null, a Bayes factor larger than 10 as "strong evidence" for the null, and a Bayes factor larger than 30 as "very strong evidence" for the null. Similarly, a Bayes factor less than 0.33 [0.10, 0.03] is counted as some [strong, very strong] evidence for the alternative hypothesis.

[6] Hence, we also defined $CE((€200, p^N), q^N) > CE(€200, r^N)$ if $CE((€200, p^N), q^N) - CE(€200, r^N) > 2$ and $CE((€200, p^N), q^N) < CE(€200, r^N)$ if $CE((€200, p^N), q^N) - CE(€200, r^N) < -2$.

[7] In the consistency tests and the tests of reduction of compound gambles that we report in Section 3.4 we used Bayesian t-tests. In these tests we did not have to use interpolation and a substantial proportion of the subjects stated the same certainty equivalents. Using tests of proportions here would make the analysis less informative and underestimate the support for the null hypothesis.

In the individual subject analyses, we classified individual subjects based on the number of times they displayed the patterns $((€200, p^N), q^N) - CE(€200, r^N) < -2$, $-2 < CE((€200, p^N), q^N) - CE(€200, r^N) < 2$, and $CE((€200, p^N), q^N) - CE(€200, r^N) > 2$ for both $N = 2$ and $N = 3$. For 2-reduction invariance, we defined subjects who reported $CE((€200, p^2), q^2) - CE(€200, r^2) < -2$ more than twice as *Type compound < simple*. We only required them to display this pattern in a majority of tests to account for response error. Similarly, we defined subjects who reported $CE((€200, p^2), q^2) - CE(€200, r^2) > 2$ more than twice as *Type compound > simple*. The other subjects were assumed to behave in line with reduction invariance (plus some error) and were defined as *Type RI*. The classification for $N = 3$ was identical.

In the individual analyses of reduction of compound gambles we defined subjects who reported $CE((€200, p), q) - CE(€200, pq) < -2$ in a majority of tests (more than 6 times) as *Type compound < simple*. Subjects who reported $CE((€200, p), q) - CE(€200, pq) > 2$ more than 6 times were defined as *Type compound > simple* and the other subjects were assumed to behave in line with reduction of compound gambles plus error and were defined as *Type RCG*.

3.4 Results

We removed one subject from the analyses because her responses reflected confusion[8]. The results presented next used the responses of the remaining 78 subjects.

3.4.1 Consistency

Each subject repeated four choices and two complete elicitations. For each subject, the repeated choices were randomly selected (and hence differed across subjects) but they were always a choice that the subject had faced in the third step of the iteration procedure. Subjects made the same choice in 72.8% of the repeated choices. Reversal rates up to one third are common in the literature (Wakker et al., 1994; Stott, 2006) and we, therefore, consider our reversal rates as satisfactory, especially if we take into

[8]In several choices, she chose 0 for sure over a gamble, which gave a positive probability of €200 and could not result in a payoff less than €0.

account that subjects were usually close to indifference in the third iteration. Fifty-four subjects (69%) had one reversal at most. Six subjects (8%) had more than two reversals. We also analyzed the data without these subjects, but this led to similar results and we do not report them. The proportions of reversals were about the same in the simple gambles and in the compound gambles: 24% versus 29% and the Bayesian 95% credible intervals overlapped.

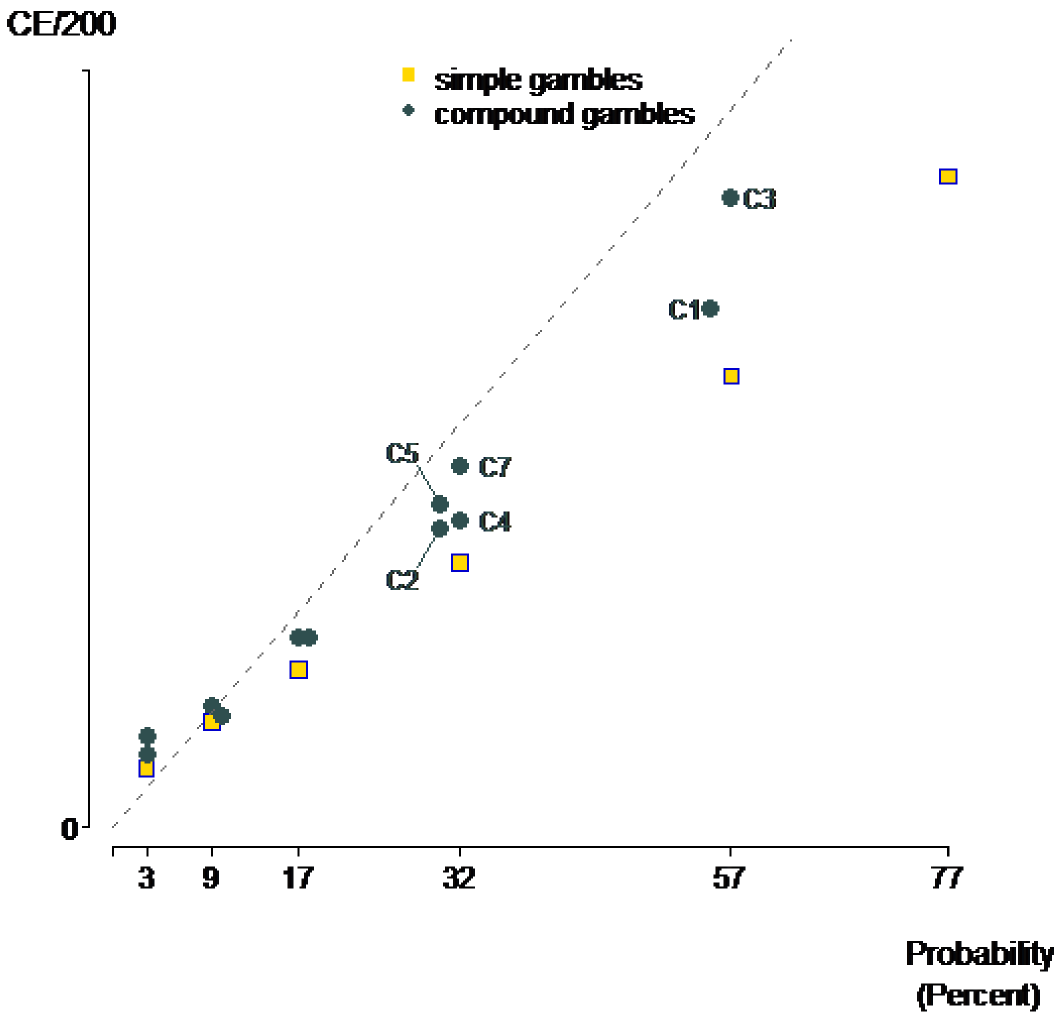
We also repeated two complete elicitations, one for a simple gamble and one for a compound gamble. Both gambles were randomly selected and, consequently, they differed across subjects. The data favored the null hypothesis of equality between the original and the repeated measurement (the Bayes factors (BFs) were 6.48 for simple gambles and 7.07 for compound gambles). The mean absolute deviation between the original and the repeated measurement was €15.38. The median was lower (€8) indicating that there were a few outliers with large differences, but for most subjects the differences were modest. The data favored the null hypothesis that the mean difference between the original and the repeated measurement was the same for the simple and for the compound gambles (BF = 7.96).

Because the questions that were repeated had different expected values, we also looked at the absolute difference as a percentage of the expected value. The mean of these percentages was 60%, the median was again much lower: 18%. The data supported the null that the means of these percentages were equal for the simple and the compound gambles (BF = 6.96) and we had no indication that subjects made more errors or had less precise preferences in the, arguably, more complex compound gambles.

3.4.2 Certainty equivalents

Figure 3.1 displays the certainty equivalents of the simple and the compound gambles. We divided these certainty equivalents by 200 to give a visual impression of subjects' risk attitudes. For risk neutral subjects, the certainty equivalents of the simple gambles (the squares in the figure) will lie on the diagonal; points above the diagonal reflect risk seeking and points below the diagonal reflect risk aversion. The figure shows the usual pattern of risk seeking for small probabilities and risk aversion for moderate and large probabilities, which is equivalent to inverse S-shaped probability weighting if utility is linear.

Figure 3.1: Mean certainty equivalents (divided by 200) of the simple and the compound gambles



3.4.3 Tests of reduction invariance

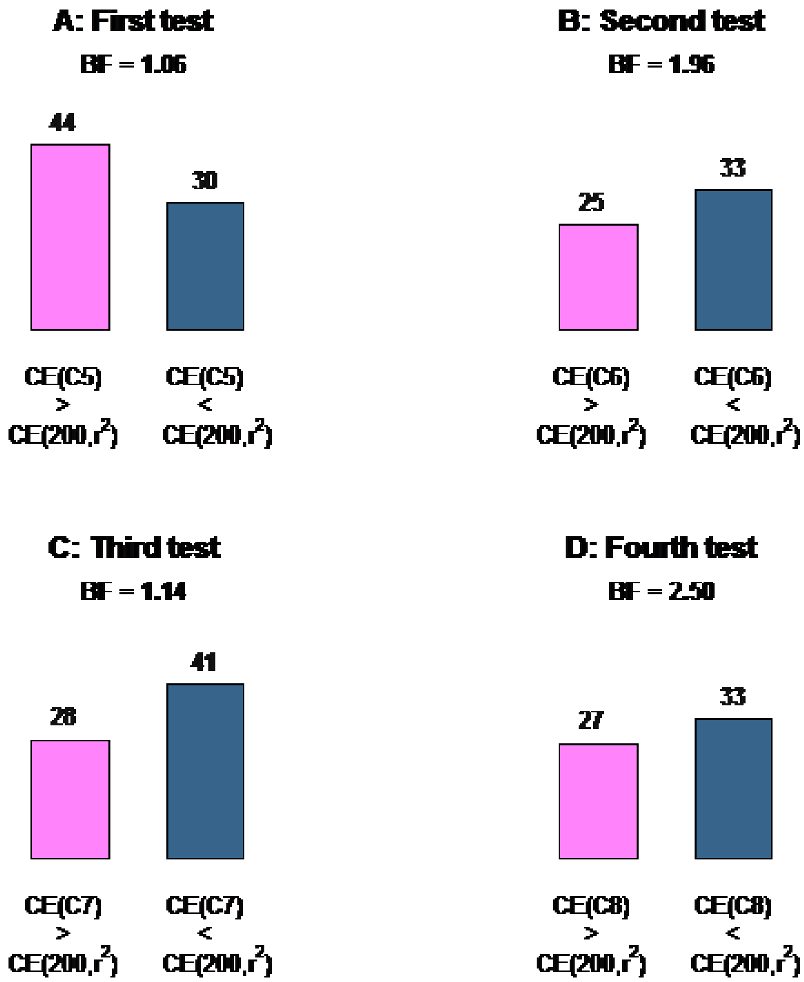
Figure 3.2 shows the results of the eight tests of reduction invariance that we performed. Panel I shows the results of the four tests of 2-reduction invariance and Panel II those of the four tests of 3-reduction invariance. For each test we have indicated the

BF-values.

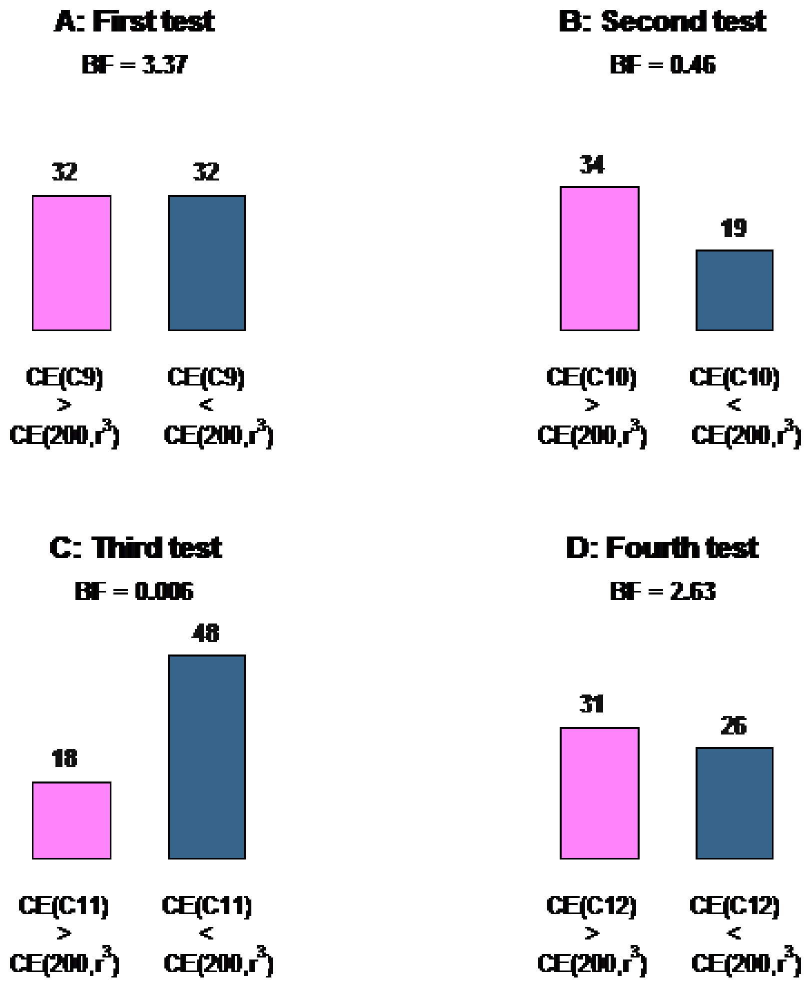
Figure 3.2: Mean certainty equivalents (divided by 200) of the simple and the compound gambles

Note: the figure shows the number of subjects for whom the certainty equivalent of the compound gamble is greater than respectively smaller than the certainty equivalent of the simple gamble (taking into account the imprecision in our measurements). Under reduction invariance these numbers should not differ systematically. BF stands for Bayes factor with higher values indicating more support for the null hypothesis that reduction invariance holds.

I: Tests of 2-reduction invariance



II: Tests of 3-reduction invariance



Pooled over all tests, the data supported the null hypothesis that reduction invariance held ($BF = 5.34$). This was also true if we look at the tests of 2-RI ($BF = 4.77$) and 3-RI ($BF = 5.12$). If we look at the eight tests separately, the data did not provide much support for either the null or the alternative. The exception was the third test of 3-RI which provided very strong evidence for the alternative that reduction invariance did not hold and the first test of 3-RI which provided some evidence for reduction invariance.

Table 3.3 shows the classification of the subjects. Reduction invariance was the dominant type with 45% of the subjects satisfying it in both tests. No other type was close to reduction invariance. Both in the tests of 2-RI and in the tests of 3-RI around

60% of the subjects satisfied reduction invariance. Two thirds of the subject could be classified the same way in both the 2-RI and the 3-RI tests. The data support the hypothesis that amongst the subjects who could be classified the same way those who behaved according to reduction invariance were more common than those who did not behave according to reduction invariance ($BF = 3.81$).

Table 3.3: Classification of subjects in the 2-reduction invariance (2-RI) and the 3-reduction invariance (3-RI) tests

	Type	**2-RI**			Total
		compound >simple	RI	compound <simple	
3-RI	compound >simple	6	8	0	14
	RI	8	35	5	48
	compound <simple	1	4	11	16
	Total	15	47	16	78

3.4.4 Tests of reduction of compound gambles

The general picture that emerges from our results is that reduction invariance was supported. This poses the question whether the special, rational case of reduction invariance, reduction of compound gambles, also held. Our results indicate that it did not hold at the aggregate level. Figure 3.1 gives a visual impression. The circles show the certainty equivalents of the compound gambles plotted against the reduced probabilities. If reduction of compound gambles held the circles and the squares should overlap. It is clear from the Figure that they did not. Bayesian tests revealed very strong evidence for the alternative hypothesis that reduction of compound gambles should be rejected ($BF = 1.14e - 23$)[9].

However, at the individual level we observed that around 47 (60%) of the subjects behaved in line with reduction of compound gambles (taking account of preference imprecision). The subjects who deviated from it, deviated overwhelmingly in the direction of higher certainty equivalents for the compound gambles than for the corresponding

[9]The pairwise tests supported the alternative hypothesis that reduction of compound gambles did not hold with Bayes factors less than .33 except for the differences between C6 and S2 ($BF = 3.14$) and between C8 and S2 ($BF = 5.70$) where the data gave some evidence for reduction of compound gambles and the differences between C11 and S3 ($BF = 0.68$), C2 and S4 ($BF = 1.07$), and C4 and S4 ($BF = 0.94$) where the data supported neither the null nor the alternative hypothesis.

simple gambles (according to the Bayes factors the posterior probability that a subject who deviated from reduction of compound gambles had a higher certainty equivalent for the compound gamble was 5642 times as high as the probability that he had a higher certainty equivalent for the corresponding simple gamble).

3.4.5 Robustness

We used linear interpolation in the analysis of reduction invariance to determine $CE(200, r^2)$ and $CE(200, r^3)$. A problem in this analysis was that we could not always determine r uniquely. We, therefore, also used interpolation by smoothing splines, a nonparametric regression technique which smoothens out response errors. The fit was good for most subjects.

The figures for this robustness check are in Appendix C. Overall, the robustness check led to the same conclusions as the analysis using linear interpolation. Based on the pooled data, the support for reduction invariance increased compared to the analysis using linear interpolation ($BF = 8.63$). The results of the separate tests were largely similar to those under linear interpolation except that in the third test of 2-RI we now also observed some evidence that reduction invariance did not hold. However, this was no longer true if we did not take imprecision into account (then the Bayes factor became 0.80). The support against reduction invariance in the third test of 3-RI decreased from very strong evidence to some evidence.

At the individual level, reduction invariance was still clearly the dominant pattern and the numbers were close to those observed under linear interpolation.

3.5 Discussion

Our data largely supported reduction invariance, the central condition underlying Prelec's (1998) compound invariant weighting function. At the aggregate level our data provided some evidence in favor of reduction invariance and at the individual level reduction invariance was clearly the dominant pattern. The only test in which we found strong evidence for the alternative hypothesis that reduction invariance did not hold was the third test of 3-RI. We do not know why this happened. The reduced probability in the third test of 3-RI was similar to that in the first test of 3-RI where we found evidence for reduction invariance. The fact that in the third test of 3-RI p was less than q cannot explain the observed violation of reduction invariance either as this was also true in, for example, the second test of 2-RI where the null of reduction invariance was supported over the alternative.

Our tests of reduction invariance require the use of measured certainty equivalents. Luce (2000) argues that certainty equivalents may lead to biased estimations of the subjective values of gambles due to inherently different attitudes towards gambles (multi-dimensional entities) and certain money amounts (one-dimensional entities). von Nitzsch and Weber (1988) demonstrated empirical evidence of this bias. This problem could be avoided by matching gambles with gambles, i.e. by directly elicitating r such that $((x,p),q) \sim (x,r)$ and then checking whether $((x,p^N),q^N) \sim (x,r^N)$,$N$=2,3. As Luce (2001) pointed out, this test carries the risk that subjects will give the salient answer $pq = r$ in spite of the many observed empirical violations of reduction of compound gambles. We, therefore, followed Luce's (2001) suggestion to use certainty equivalents in the tests of reduction invariance. To reduce possible distortions, we used a choice-based procedure to determine the certainty equivalents. Previous evidence suggests that observed anomalies are substantially reduced when choice-based certainty equivalents are used instead of judged certainty equivalents (Bostic et al., 1990; von Winterfeldt et al., 1997). The procedure we used is close to the PEST procedure used by Luce in his experimental research (Chung et al. 1994, Cho et al. 1994, Cho and Luce 1995).

We used several ways to account for the stochastic nature of people's preferences. Rather than testing equality of certainty equivalents we followed Cho and Luce (1995) and tested whether the proportion of subjects for whom $CE((200,p^N),q^N)$ exceeded $CE(200,r^N)$ was the same as the proportion of subjects for whom $CE((200,p^N),q^N)$ was less than $CE(200,r^N)$. Moreover, we accounted for the imprecision in our measurements and in the individual analyses we only required preference patterns to hold in a majority of cases. There exist different and more sophisticated procedures to model choice errors. For example, Davis-Stober (2009) derived statistical tests based on order-constrained inference techniques, which were applied, amongst others in Regenwetter et al. (2011) to test transitivity and in Davis-Stober et al. (2015) to compare models based on strict weak order representations with those based on lexicographic semiorder representations. It is interesting to repeat our analysis using these methods, but it should be realized that they are, to the best of our knowledge, not yet applicable to matching tasks and that they require each choice to be repeated many times. In our experiment subjects made around 100 choices, but if we were to use the same amount of repetitions as Regenwetter et al. (2011) or Regenwetter and Davis-Stober (2012) did, subjects would have to make more than 2000 choices, which might reduce accuracy.

We found mixed support for reduction of compound gambles, the rational special case of reduction invariance. The condition was clearly violated at the aggregate level, but 60% of the subjects behaved in line with it. The violations of reduction of compound gambles that we observed indicate that subjects generally preferred compound gambles to simple gambles giving the same reduced probability. This compound risk seeking is consistent with Friedman (2005) and Kahn and Sarin (1988). It could be explained by a utility of gambling (Luce and Marley, 2000; Luce et al., 2008) as the compound gambles offer the possibility to gamble twice. On the other hand, Abdellaoui et al. (2015) observed that their subjects were compound risk averse and preferred simple gambles with the same reduced probability. They also observed that subjects became more compound risk averse for higher probabilities, while we observed the opposite pattern. The range of probabilities Abdellaoui et al. explored is larger than the range we explored. Moreover, the compound gambles for which they found compound risk aversion were more complex than the compound gambles we used and it was more difficult for their subjects to compute the reduced probabilities. Complexity aversion may have contributed to compound risk aversion in their study.

We obtained some evidence that when choosing between two gambles with the same expected value, subjects preferred the gamble with the higher second-stage probability to the gamble with the higher first-stage probability. This is consistent with a preference to have most uncertainty resolved at the first stage and violates event commutativity (Luce 2000). We found very strong evidence that the certainty equivalent of $C7$, which offered a higher probability at the second stage, was higher than the certainty equivalent of $C4$, which offered the approximately the same reduced probability but a higher first-stage probability (according to the Bayes factors, the posterior probability that $CE(C7) > CE(C4)$ was 471 times as high as the probability that $CE(C7) < CE(C4)$). More support for a preference to have the high probability resolved later comes from a comparison of compound gambles $C1$ and $C3$, which were also close in reduced probability. We found very strong evidence that the certainty equivalent of $C3$, which offered a larger second-stage probability exceeded that of $C1$, which offered a larger first-stage probability (odds 56.93). On the other hand, we also found strong evidence that the certainty equivalent of gamble $C5$ exceeded the certainty equivalent of gamble $C2$ (odds 20.41), which is inconsistent with a preference to have the high probability resolved later. As mentioned above, Ronen (1973) and Budescu and Fischer (2001) obtained clear evidence to have the high probability resolved first. Budescu and Fischer (2001) observed that hope was an important reason why their subjects preferred higher initial probabilities. A typical reason subjects gave was that "the progress from one stage to the other means something, it's better to lose at a later stage". Apparently, such considerations played no role in our study or they were offset by other considerations such as disappointment aversion which predicts that the high probability will be resolved later.

3.6 Conclusion

Prelec's (1998) compound-invariant family provides a simple way to model deviations from expected utility. It has a preference foundation, its parameters are intuitive, and it has often been used in empirical research. Luce (2001) gave an elegant simplification of Prelec's central condition and our study showed evidence in support of Luce's central condition, reduction invariance. This implies that Prelec's function provides an accurate description of the way people weight probabilities and endorses its use in empirical research. Reduction of compound gambles, a special case of reduction invariance, which is often considered rational, was rejected at the aggregate level, even though 60% of the subjects behaved in line with it implying that the power probability weighting function, which depends on reduction of compound gambles, should be used with caution.

3.7 Appendix A. Instructions and comprehension questions

Welcome!

During this experiment, you will face different choice situations involving risk. In each situation, you are asked to choose between two prospects:

• Prospect A gives you an amount of money contingent on the color of a ball drawn from an urn.

• Prospect B gives you an amount of money for sure.

The outcome of Prospect A can depend on a single draw from an urn or on two consecutive draws from two different urns. Figure 3.3 shows an example of the first scenario where the outcomes of Prospect A is determined by a single draw from an urn.

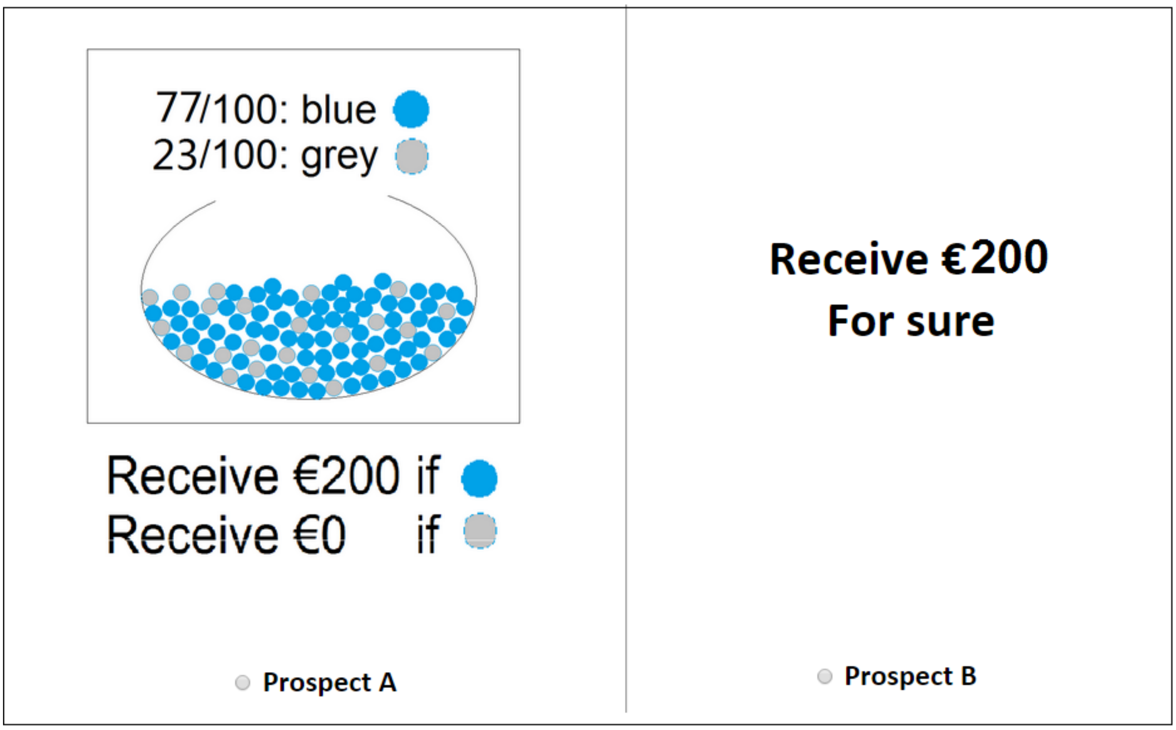


Figure 3.3: Screenshot of a question

In this choice situation, there are 100 balls in the urn, of which 77 are blue, and 23 are grey. If the drawn ball is blue, you receive €200; if it is grey, you receive €0. On the other hand, Prospect B gives you €200 for sure. In this example, you would prefer

Prospect B, because it gives you €200 for sure whereas receiving the same amount is not certain in Prospect A. Figure 3.4 presents an example of the second scenario where the outcomes of Prospect A depends on two draws.

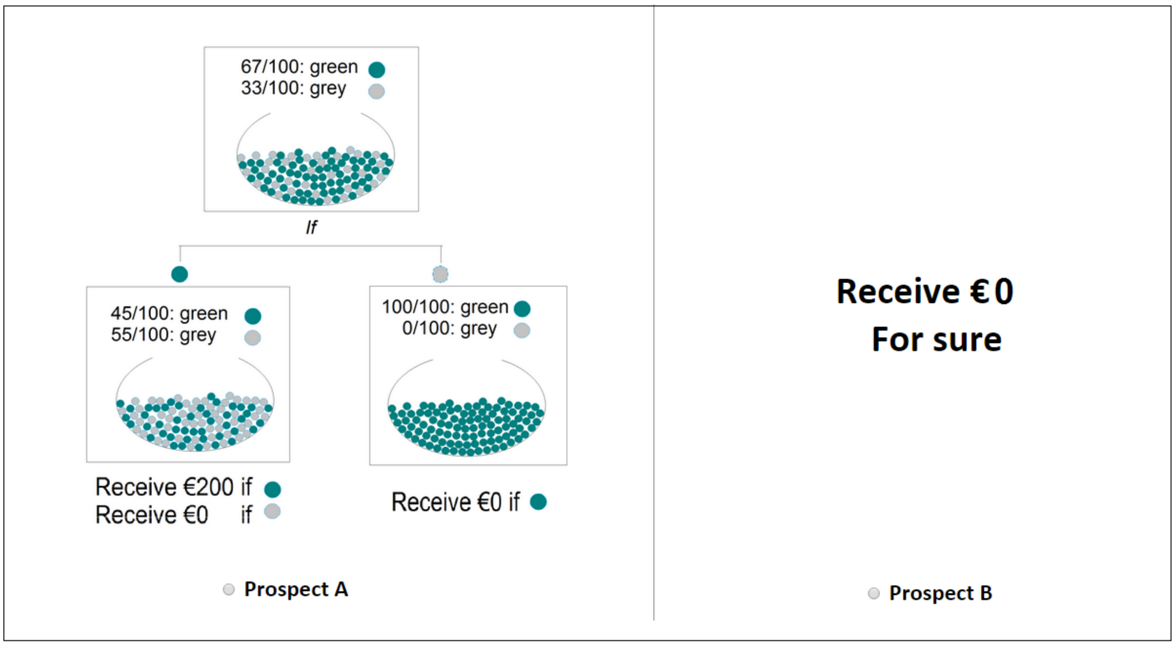


Figure 3.4: Screenshot of a question

In this choice situation, the first draw is made from the urn displayed on the top which contains 67 green and 33 grey balls. If the ball is green, then a second ball will be drawn from the left urn below; otherwise it will be drawn from the right urn below.

The final outcome will be determined by the second ball. For instance, if the second ball is drawn from the left urn, then a green ball will result in €200, and a grey ball will result in €0. If the second ball is drawn from the right urn, then the outcome will be €0 for sure because all balls in the right urn are green. On the other hand, Prospect B gives you €0 for sure. In this example, you would prefer Prospect A, because it gives you a positive chance of receiving €200 whereas Prospect B gives you €0 for sure. Once you have made your choice between two prospects, a confirm button will appear. If you agree with your choice, please click on it to go to the next question. You will not be able to change your choice after you click on the "Confirm" button.

Payment

To thank for your participation, you will receive a €5 show-up fee. In addition, two participants in this room will play out one of her choices for real. They will be selected randomly at the end of the experiment. For each of the selected participants, one of the choice situations that she faced during the experiment will be randomly selected, and his/her choice in that choice situation will be played for real.

We will now test your understanding of the instructions. Assume that you have been selected as one of the two participants who can play a question for real and that the question below was randomly selected.

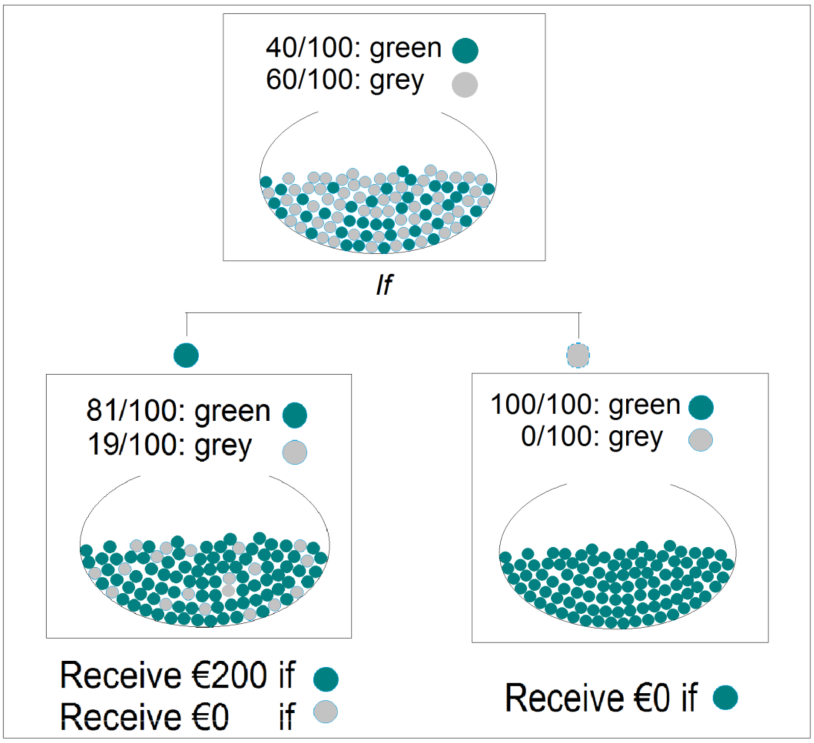


Figure 3.5: Comprehension test

Please answer the following questions.

Question 1 How many balls are there in each urn?

- 40
- 60
- 100

- 81

Question 2 In the top urn, how many green balls are there?

- 40
- 60
- 100
- 81

Question 3 In which case will you receive €200?

- Draw a grey ball in the top urn, OR draw a green ball in the bottom left urn.
- Draw a green ball in the top urn, OR draw a green ball in the bottom left urn.
- Draw a grey ball in the top urn, AND draw a green ball in the bottom left urn.
- Draw a green ball in the top urn, AND draw a green ball in the bottom left urn.

3.8 Appendix B. The iteration procedure

Subjects always chose between a gamble and a sure amount x.

1. The initial value of x was the even number closest to the expected value of the gamble.

2. x was decreased when it was chosen over the gamble and increased when the lottery was chosen.

3. The initial step size was 4, 8, 16, or 32. By choosing powers of 2 we ensured that subsequent changes were also integers. The initial step size was the number in the set 4,8,16,32 that was closest to half the initial value.

4. The step size remained constant until the subjects switched. Then it was halved.

5. The minimum step size was 2. The switching point was the midpoint between the largest value of x for which the gamble was preferred and the smallest value of x for which x was preferred.

6. If a subject had to choose between 200 for sure and the gamble or between 0 for sure and the gamble. If subjects chose the dominated option, a warning message appeared: "Please reconsider your choice". The subject was asked to choose again. If the subject continues to choose the dominated choice, we proceeded to the next elicitation.

Table 3.4 shows the initial values and the initial step sizes for the eighteen gambles in the experiment.

Table 3.4: Initial values and initial step sizes for the gambles in the experiment

Gamble	**Expected value**	**Initial value**	**Initial step size**
C1	109.88	110	32
C2	60.3	60	32
C3	113.4	114	32
C4	63.96	64	32
C5	60.3	60	32
C6	18	18	8
C7	64.8	64	32
C8	20.1	20	8
C9	33	32	16
C10	5.4	6	4
C11	36.5	36	16
C12	6.6	6	4
S1	6	6	4
S2	18	18	8
S3	34	34	16
S4	64	64	32
S5	114	114	32
S6	154	154	32

3.9 Appendix C. Tests of reduction invariance under fitting of the certainty equivalents by smoothing splines

Figure 3.6: Tests of 2-reduction invariance

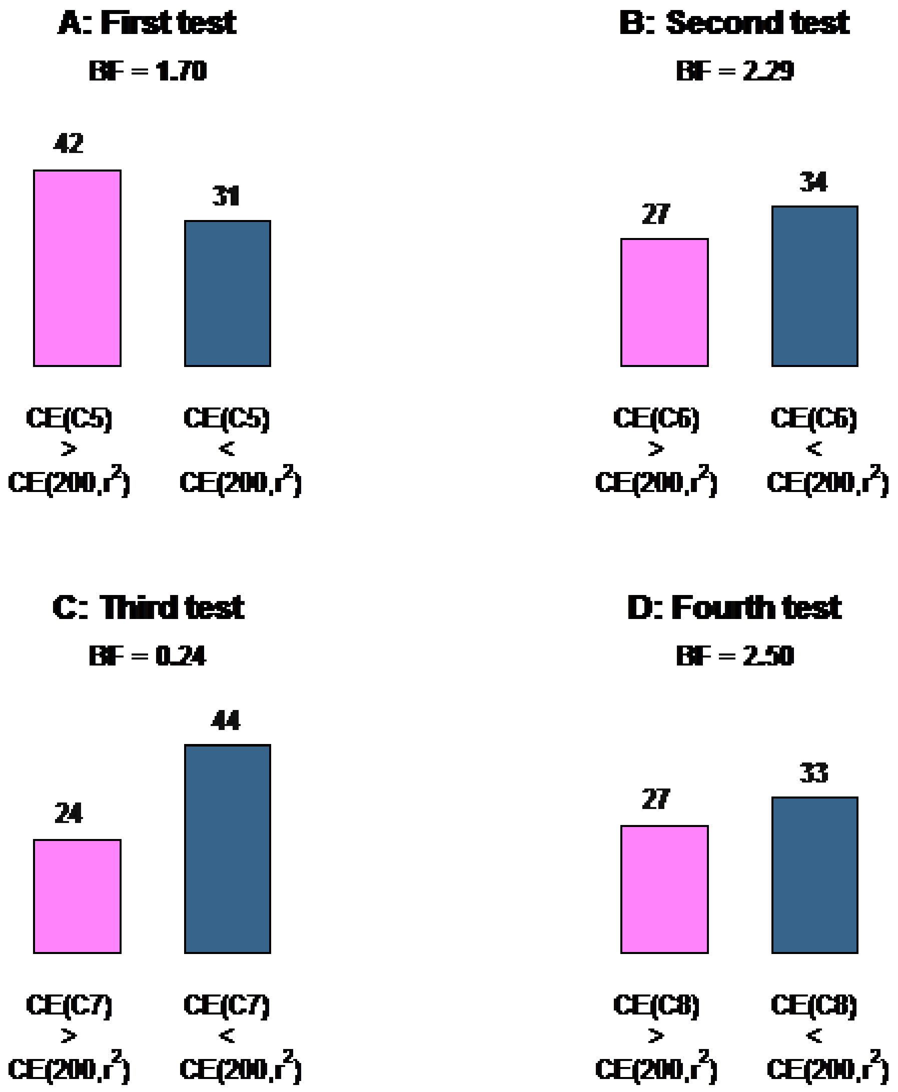


Figure 3.7: Tests of 3-reduction invariance

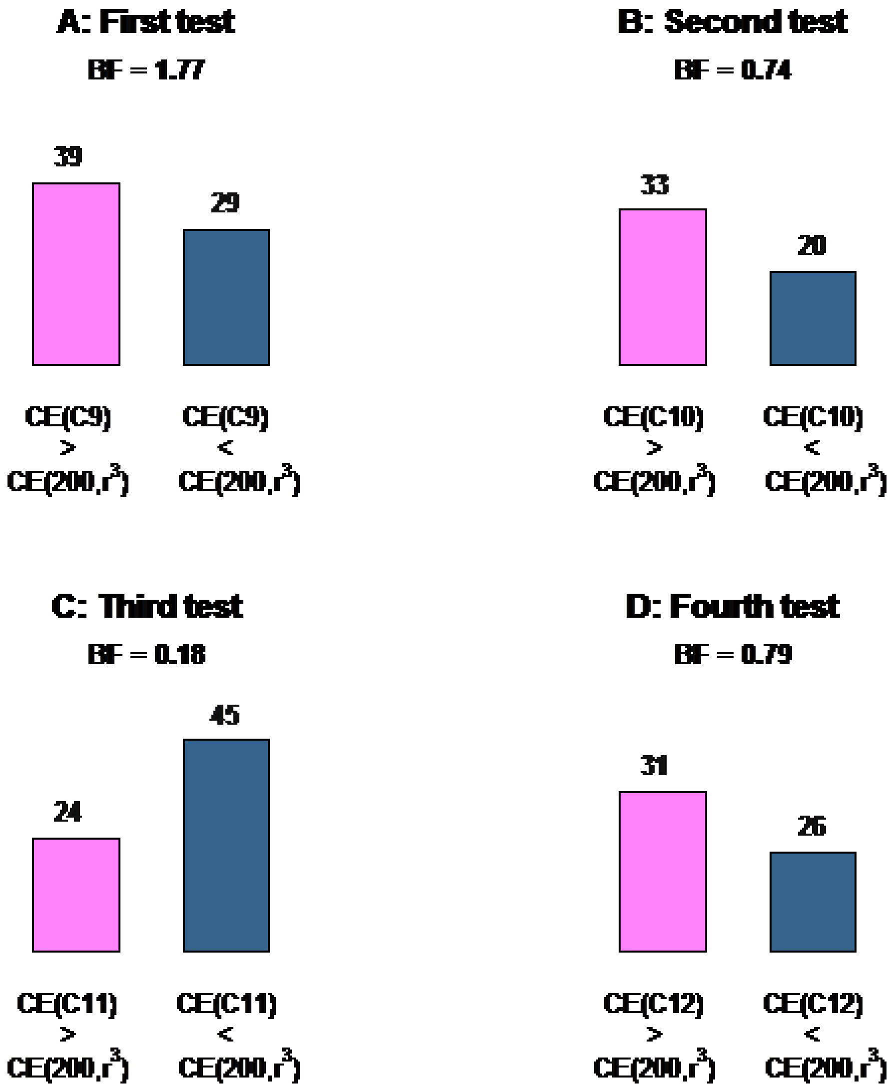


Table 3.5: Classification of subjects in the 2-reduction invariance (2-RI) and the 3-reduction invariance (3-RI) tests

Type		**2-RI**			Total
		compound >simple	RI	compound <simple	
3-RI	compound >simple	8	6	1	15
	RI	7	30	10	47
	compound <simple	1	4	11	16
Total		16	40	22	78

Chapter 4

A Measurement of Decreasing Impatience for Health and Money

Private and policy decisions often involve outcomes that occur at different points in time. Examples are choosing a pension plan and funding a screening program that reduces future illness. To account for their differences in timing, outcomes are usually discounted at a constant rate.

Constant discounting is tractable and has normative appeal, but it is inconsistent with observed behavior. Empirical evidence shows that discount rates typically decrease over time (Frederick et al., 2002; Attema, 2012). Most evidence for decreasing impatience comes from studies using money outcomes, but it has also been observed for other domains such as health and environmental outcomes (Bleichrodt and Johannesson, 2001; van der Pol and Cairns, 2011, 2002; Khwaja et al., 2007; Hardisty and Weber, 2009; Cairns and Van der Pol, 1997).

The violations of constant discounting have implications for policy. From Strotz (1955), we know that a decision maker who deviates from constant discounting may be prone to behave inconsistently over time and may have self-control problems, which lead to self-harming behaviors such as saving too little, addiction (Gruber and Köszegi, 2001) and obesity (Scharff, 2009; Ikeda et al., 2010). These self-control problems, in

This chapter is based on the homonymous paper, co-authored with Han Bleichrodt and Kirsten I.M. Rohde.

turn, may increase the welfare benefits from policy. For example, the net benefit of an increased tax on smoking may be much larger when the smoker does not discount at a constant rate, because the tax can serve as a commitment device that reduces the smoker's self-control problems and which he, therefore, values (Gruber and Köszegi, 2004). To assess the severity of departure from constant impatience, and, consequently, the vulnerability to self-control problems and the potential benefits from policy, the degree of decreasing impatience must be quantified. This is the aim of our paper.

Prelec (2004) showed that decreasing impatience cannot be quantified by looking at the speed of decline of discount rates. Hence, the above mentioned support for decreasing impatience, which compared discount rates, cannot be used to quantify decreasing impatience. Prelec argued that decreasing impatience should be measured by the Pratt-Arrow convexity of the logarithm of the discount function. Unfortunately, this measure is hard to observe empirically. Instead, we will use the method of Attema et al. (2010), which is informationally equivalent to Prelec's measure and can easily be applied empirically to measure the degree of decreasing impatience. Attema et al.'s measure makes no assumptions about utility or intertemporal separability. Existing studies on time preference generally imposed parametric assumptions on utility (most studies assumed linear utility) and assumed intertemporal separability. These assumptions cause distortions in the measurement of time preferences (Attema, 2012; Broome, 1991; Loewenstein and Prelec, 1993). Finally, Attema et al.'s method allows analyses at the individual level and, as we will show, individual time preferences are heterogeneous.

In an experiment, we compared deviations from constant discounting for money and health, two domains where economic analyses are widely used and discounting is routinely applied. Knowing whether time preferences are the same for health and money is important for both research and policy. Researchers often assume the same (constant) discounting of money and health and government offices try to set a single official discount rate to evaluate all public investments. If capital markets worked perfectly then this would be the appropriate discounting policy (Moore and Viscusi, 1990). However, health is less transferable over time than money and there is no market for health to observe. In the presence of such market imperfections, it is unclear whether health and money should be discounted similarly. As noted by Moore and Viscusi (1990, p.52), this question must be resolved empirically, which is what the current paper seeks to do.

Several papers have compared discount rates for health and money. As mentioned above, the results from these studies cannot answer whether the degree of decreasing impatience differs between health and money, but they do shed light on whether people discount health and money similarly. The results are mixed (Attema 2012). While Moore and Viscusi (1990) and Cropper et al. (1994) found the same discounting for health and money, Cairns (1992) found more discounting for money and Cairns (1994) and Hardisty and Weber (2009) found more discounting for health gains and less for health losses. Moreover, the correlation between discounting for health and discounting for money is typically low (Chapman and Elstein, 1995; Chapman, 1996a).

The empirical deficiencies of constant discounting have led to a variety of new discount models. The most widely-used of these is quasi-hyperbolic discounting (Phelps and Pollak, 1968; Laibson, 1997), which has become part of mainstream economics (Gruber and Köszegi, 2001; Diamond and Köszegi, 2003; DellaVigna, 2009). Empirical evidence on the relative performance of these new discount models is thin on the ground, especially for health. This is unfortunate given the increasing use of these models in health (Gruber and Köszegi, 2001, 2004; Newhouse, 2006; Fang and Wang, 2015). A final contribution of this paper is to present evidence about the descriptive validity of discount models.

Our results indicate that most subjects deviated from constant discounting and were decreasingly impatient for both money and health. Between 25% (for health) and 35% (for money) of our subjects behaved according to increasing impatience, a finding that most discount models cannot explain. Subjects deviated more from constant discounting for health than for money. This domain-dependence of discounting suggests that evidence on time preferences for money has only limited validity for health. Of the discounting models that we explored, hyperbolic discounting (Loewenstein and Prelec, 1992) and proportional discounting (Mazur, 1987) described time preferences for health and money best. Quasi-hyperbolic discounting and constant discounting could be rejected for both health and money.

4.1 Background

We consider a decision maker's preferences $\succcurlyeq$ over *timed outcomes (x,t)*, which denotes "receiving outcome x at time t". Outcomes are health states or money amounts in our experiment. Time point *t=0* is the present. We denote strict preference by $\succ$, indifference by $\sim$, and reversed preferences by $\preccurlyeq$ (weak reversed preference) and $\prec$ (strict reversed preference). Throughout the paper, we assume that the decision maker evaluates timed outcomes using *discounted utility*:

$$DU(x,t) = \phi(t)U(x). \tag{4.1}$$

In Eq.(4.1), ϕ is a decreasing and positive *discount function* and U is a real-valued *utility function*. Because ϕ is decreasing, the decision maker is *impatient* and prefers to receive good outcomes sooner rather than later. We scale ϕ such that $\phi(0) = 1$. Utility is defined relative to a neutral outcome which has the value 0. For money the neutral outcome was receiving nothing, for health we selected a specific health state (chronic back pain) that we assigned the value 0.

Constant impatience says that preferences between timed outcomes do not change if we delay them by a common constant: for all $\sigma > 0$, $(x, s) \sim (y, t)$ with $0 \prec x \prec y$ and $s < t$ implies $(x, s+\sigma) \sim (y, t+\sigma)$. Koopmans (1960) showed that constant impatience implies constant discounting: $\phi(t) = \sigma^t$ for $0 < \sigma < 1$. *Decreasing impatience* holds if adding a common delay makes people more willing to wait for the better outcome: for all $\sigma > 0$,$(x, s) \sim (y, t)$ with $0 \prec x \prec y$ and $s < t$ implies $(x, s + \sigma) \preccurlyeq (y, t + \sigma)$. Empirical studies have often found decreasing impatience, for both money (Frederick et al., 2002) and health (Attema, 2012). *Increasing impatience* is the opposite of decreasing impatience and means that adding a common delay makes people less willing to wait for a larger outcome: for all $\sigma > 0$,$(x, s) \sim (y, t)$ with $0 \prec x \prec y$ and s<t implies $(x, s+\sigma) \succcurlyeq (y, t+\sigma)$. Several studies have found increasing impatience for money (e.g. Attema et al., 2010; Sayman and Öncüler, 2009; Scholten and Read, 2006; Loewenstein, 1987; Takeuchi, 2011). For health, only indirect evidence of increasing impatience exists (Attema et al., 2012).

Let $\succcurlyeq_a$ and $\succcurlyeq_b$ be the preference relations over timed outcomes of decision makers a and b. We say that $\succcurlyeq_b$ is more decreasingly impatient than $\succcurlyeq_a$ if for all $0 < s < t$, for all ϵ, and for all outcomes $0 \prec_a x \prec_a y, 0 \prec_b x' \prec_b y'$, if $(x,s) \sim_a (y,t), (x, s+\sigma) \sim_a (y, t+\sigma+\epsilon)$ and $(x', s) \sim_b (y', t)$ then $(x', s+\sigma) \prec_b (y', t+\sigma+\epsilon)$. Intuitively, if both a and b are willing to wait from period s to period t to receive a larger outcome (y instead of x for decision maker a and y' instead of x' for decision maker b), they are equally impatient for these outcomes and periods. Now, if a is also willing to wait from period $s+\sigma$ to period $t+\sigma+\epsilon$ to receive y instead of x then b will prefer the larger later outcome $(y', t+\sigma+\epsilon)$, because his impatience decreases faster than that of a and, thus, he becomes more future-oriented than a.

Analogously, $\prec_b$ is more *increasingly impatient* than $\prec_a$ if for all $0 < s < t$, for all $\sigma > 0$, for all ϵ, and for all outcomes $0 \prec_a x \prec_b y, 0 \prec_b x' \prec_b y'$, if $(x,s) \sim_a (y,t)$, $(x, s+\sigma) \sim_a (y, t+\sigma+\epsilon)$ and $(x', s) \sim_b (y', t)$ then $(x', s+\sigma) \succ_b (y', t+\sigma+\epsilon)$.

Various alternative models have been proposed to accommodate the deviations from constant discounting. The most popular of these models is *quasi-hyperbolic discounting* (Phelps and Pollak, 1968; Laibson, 1997):

$$\phi(t) = \begin{cases} \beta\sigma^t \; for \; t > 0 \\ 1 \; for \; t = 0. \end{cases} \tag{4.2}$$

with $0 < \beta, \sigma < 1$. Quasi-hyperbolic discounting differs from constant discounting only in the first period. The model assumes that a decision maker gives extra weight to the present and the parameter β captures this *present bias.* Present bias leads to decreasing impatience in the first period and constant impatience in all later periods.

Loewenstein and Prelec (1992) proposed a more general model, *hyperbolic discounting*, in which decreasing impatience not only occurs in the first period but also in later periods:

$$\phi(t) = (1+ht)^{(-r/h)}, h \geq 0, r > 0. \tag{4.3}$$

The parameter h measures decreasing impatience. If h=*0* then hyperbolic discounting is equivalent to constant discounting and the larger is h the more the decision maker deviates from constant discounting. Two special cases of hyperbolic discounting are proportional discounting (Mazur, 1987), which results from Eq.(4.3) when h=r and power discounting (Harvey, 1986), which results when h=*1*.

For money, Abdellaoui et al. (2013, 2010) concluded that hyperbolic discounting performed better than constant, quasi-hyperbolic, proportional, and power discounting, even after correction for the differences in degrees of freedom. For health, van der Pol and Cairns (2002) found some evidence that hyperbolic discounting and power discounting fitted better than constant discounting and proportional discounting. Bleichrodt and Johannesson (2001) and van der Pol and Cairns (2011) found that hyperbolic discounting fitted better than constant discounting and quasi-hyperbolic discounting for health.

4.2 Time trade-off sequences

Because deviations from constant impatience are closely related to economic and health misbehaviors, it is of interest to measure these deviations[1]. Prelec (2004) argued that deviations from constant impatience should be measured by the Pratt-Arrow convexity of the logarithm of the discount function:$-ln(\phi)''/ln(\phi)'$. This measure is hard to observe empirically. First, we must measure the discount function, which is complex because discounting and utility interact, then we must take the logarithm, and, finally, we must compute first and second derivatives.

Attema et al. (2010) showed that deviations from constant impatience can be measured more easily using time trade-off sequences. To illustrate, we first choose two outcomes x and y with $x \prec y$. A time trade-off sequence is a sequence of time points t_0,t_1,...,t_k such that

$$\begin{gathered}(x, t_0) \sim (y, t_1) \\ (x, t_1) \sim (y, t_2) \\ \vdots \\ (x, t_{k-1}) \sim (y, t_k)\end{gathered} \tag{4.4}$$

We call $WTW_i = t_i - t_{i-1}, i = 1, ..., k$, the decision maker's willingness to wait. Con-

[1]Strictly speaking, violations of constant impatience are not equivalent to time inconsistent behavior (reversals of preference over time) and self-control problems (Harvey, 1995). However, they are equivalent under the common assumption of time invariance (Halevy, 2015).

stant impatience implies that the willingness to wait is constant, decreasing impatience implies that the willingness to wait increases with i, and increasing impatience implies that the willingness to wait decreases with i. From Eq.(4.1), we obtain

$$\phi(t_0)/\phi(t_1) = \phi(t_1)/\phi(t_2) = ... = \phi(t_{k-1})/\phi(t_k) \tag{4.5}$$

This is equivalent to:

$$ln(\phi(t_0)) - ln(\phi(t_1)) = ln(\phi(t_1)) - ln(\phi(t_2)) = ... = ln(\phi(t_{k-1}) - ln(\phi(t_k)) \tag{4.6}$$

Eq.(4.6) shows that a time trade-off sequence is equally spaced in terms of $ln(\phi)$. This property does not depend on utility. Utility drops from Eqs.(4.5) and (4.6) and we do not have to make any assumptions about it.

We now define the *time curve*

$$\tau(t) = \frac{ln(\phi(t)) - ln(\phi(t_k))}{ln(\phi(t_0)) - ln(\phi(t_k))} \tag{4.7}$$

From Eq.(4.7),$\tau(t_0) = 1$, $\tau(t_k) = 0$, and $\tau(t_j) = 1 - j/k$. Because $\tau(t_j) = 1 - j/k$, the elements of the time trade-off sequence are also equally spaced in terms of τ. Under constant discounting τ is linear, under decreasing impatience it is convex, and under increasing impatience it is concave. Attema et al. (2010) showed that τ has the same degree of convexity as $ln(\phi) = -\tau''/\tau' = -ln(\phi)''/ln(\phi)'$. In other words, τ can be used instead of $ln(\phi)$ to measure decreasing impatience and decision maker a is more decreasingly impatient than decision maker b if a's time curve is more convex than b's time curve. The big advantage of using τ instead of $ln(\phi)$ is that τ is directly observable whereas $ln(\phi)$ is not.

The time curve can also be used to test the different discount models. Rohde (2010) proposed the *hyperbolic factor* :

$$hyp(i, j) = \frac{(t_j - t_i) - (t_{j-i} - t_{i-1})}{t_i(t_{j-1} - t_{i-1}) - t_{i-1}(t_j - t_i)} \tag{4.8}$$

with $t_i < t_j$ and derived that:

Observation 1 (Rohde 2010): The hyperbolic factor is:

1. equal to zero under constant discounting,
2. positive if $t_{i-1} = 0$ and zero if $t_{i-1} > 0$ under quasi-hyperbolic discounting,
3. equal to $h > 0$ under hyperbolic discounting. Moreover, under hyperbolic discounting the denominator of Eq.(4.8) should be positive,
4. equal to r under proportional discounting, and
5. equal to 1 under power discounting.

4.3 Experiment

Our experiment elicited time trade-off sequences for health and money. We recruited seventy-five students (36 female) from Erasmus University Rotterdam, mainly from economics and business. Every subject received a €12 participation fee. The experiment was computer-run in 14 small group sessions. Subjects were seated in cubicles and could not see each other's screens or interact.

The experiment consisted of two parts, the elicitation of the time trade-off sequences for health and for money. We randomized the order of these parts. Each part started with instructions and four comprehension questions (see the online appendix). After a subject had correctly answered all four comprehension questions, he answered two training questions. We told subjects that the training questions and the experimental questions had no right or wrong answers and that we were only interested in their preferences. Subjects were encouraged to ask questions at any time they wished should something be unclear.

Table 4.1: Stimuli of the four sequences

Parts	**Sequence**	t_0	x	y
Health	**H1**	immediately	Treatment A	Treatment B
	H2	4 weeks	Treatment A	Treatment B
Money	**M1**	Immediately	500 euro	550 euro
	M2	4 weeks	500 euro	550 euro

We measured four time trade-off sequences for each subject, two for health and two for money. Table 4.1 shows the stimuli. All delays were in weeks. For both health and money, one sequence started immediately and the other in four weeks. We randomized which of these sequences came first.

For money we used $x =$ €500 and $y =$ €550 to elicit the time trade-off sequences. For health, we told subjects to imagine that they suffered from chronic back pain (the neutral level). Chronic back pain was described as:

- You have moderate problems in walking about.
- You have moderate problems performing your usual activities
(e.g., work, study, housework, family or leisure activities).
- You have moderate pain or discomfort.

We told subjects that there are two treatments (A and B) to relieve chronic back pain. Table 4.2 shows the descriptions of the two treatments, which were presented to subjects on their computer screens. For easy reference, they were also printed on cards, which we put on subjects' desks. Treatment B was more effective than Treatment A. Both treatments removed the pain, but B also improved walking and the performance of usual activities. The effects of the treatments start immediately at the beginning of the treatment and last for exactly one week, the unit of time we use in this paper. After this week, chronic back pain returns. Such questions are common in the measurement of time preferences for health (e.g. Van der Pol and Cairns 2011, Hardisty and Weber 2009) except that subjects usually consider only one change in health (e.g. Treatment A) and the duration of this change is varied. There are two advantages of keeping the duration of change fixed. First, the utility for time duration can be entirely general. Studies that vary the duration of change have to impose simplifying assumptions on the utility for time duration to be able to analyze the responses and most studies assume it is linear. A second advantage of keeping the duration of change fixed is that subjects will more likely concentrate on the time point at which the change occurs, which is desirable as we are interested in the properties of the discount function and not in those of the utility function. Our instructions told subjects to adopt chronic back pain as their neutral level of health. Because most subjects were healthy, chronic back pain could have been perceived as a loss and not as neutral. However, empirical evidence suggests that the reference point or neutral level of health can be manipulated and even healthy subjects usually adopt a health state which is worse than their current health if instructed to do so (Bleichrodt and Pinto, 2002; Attema et al., 2013; van Osch et al., 2006).

Table 4.2: The descriptions of the treatments

Treatment A	Treatment B
During one week · You have **moderate problems** in walking about. · You have **moderate problems** with performing your usual activities (e.g. work, study, housework, family or leisure activities). · You have **no pain** or **discomfort.**	*During one week* · You have **slight problems** in walking about. · You have **no problem** with performing your usual activities (e.g. work, study, housework, family or leisure activities). · You have **no pain** or **discomfort.**

Each sequence consisted of four elements ($k = 4$). All indifferences were elicited using a series of choices. This procedure is common in experimental economics, because it leads to fewer inconsistencies than directly asking subjects for their indifference values (Bostic et al., 1990). We will describe the choice-based elicitation procedure for health. It was similar for money with €500 instead of Treatment A and with €550 instead of Treatment B. Subjects first made several pairwise choices. These choices limited the range within which their willingness to wait fell. Figure 4.1 gives an example of a pairwise choice for health.

Figure 4.1: An example of a pairwise choice for health

Option A	A	B	Option B
Treatment A immediately	○	○	Treatment B in 50 weeks

In the first pairwise choice, the benefits of Treatment B occurred in 100 weeks. The next choices then zoomed in on subjects' willingness to wait. Once the range within which their willingness to wait fell had been narrowed to at most 13 weeks, subjects filled out a choice list. Figure 4.2 gives an example. The first and final choice on the list had been made before. So in Figure 4.2 the subject had already chosen B in 12 weeks over A immediately and A immediately over B in 25 weeks. Consistency requires that a subject switches from B to A at some choice in the list. If the subject always chose the same treatment then the elicitation would recommence for this question starting with the first pairwise choice. If the subject was also inconsistent in the repeated elicitation then we treated the response to this question as missing. There were six subjects who never switched in at least one choice list.

Figure 4.2: An example of a choice list for health

Option A	A	B	Option B
Treatment A immediately	○	○	Treatment B in 12 weeks
	○	○	Treatment B in 13 weeks
	○	○	Treatment B in 14 weeks
	○	○	Treatment B in 15 weeks
	○	○	Treatment B in 16 weeks
	○	○	Treatment B in 17 weeks
	○	○	Treatment B in 18 weeks
	○	○	Treatment B in 19 weeks
	○	○	Treatment B in 20 weeks
	○	○	Treatment B in 21 weeks
	○	○	Treatment B in 22 weeks
	○	○	Treatment B in 23 weeks
	○	○	Treatment B in 24 weeks
	○	○	Treatment B in 25 weeks

The upper bound for the delay in Option B was 500 weeks. If a subject still preferred B for a delay of 500 weeks, we also treated his response to this question as missing. Six subjects did this at least once. These subjects were the most patient. To test whether the exclusion of the most patient subjects biased the results, we repeated the individual analyses by also excluding the six most impatient subjects. This robustness check led to the same conclusions in all but one case and we will only report the single case where the results differed.

4.4 Results

We removed a subject's missing choices, but kept the other, completed, choices in the aggregate analyses. In the individual analyses, we needed all choices and the 12 subjects with missing data were removed[2]. The individual analyses, therefore, used the responses of 63 subjects.

[2]Six subjects who never switched between Options A and B and eight who were extremely patient. Two of the extremely patient subjects never switched either. Therefore we excluded 12 subjects in total.

4.4.1 Consistency

For each subject we repeated two, randomly selected, elicitations, one for health and one for money to assess data quality. The consistency of the measurements was good. The original and the repeated measurements did not differ, neither for health (Wilcoxon test, $p = 0.86$) nor for money (Wilcoxon test, $p = 0.58$). The median absolute difference between the original and the repeated measurement was one week for both health and money.

4.4.2 Aggregate results

Figure 4.3 shows the four time curves based on the mean data. The figures based on the median data are similar. The dashed lines correspond to constant discounting. For health, Panels A and B show that the mean data deviated from constant discounting. The convex shape of the time curves indicates that subjects were decreasingly impatient[3]. We could reject the null hypothesis of constant impatience against the alternative of decreasing impatience in both sequences (Page's L-test test, both $p < 0.01$). The data are inconsistent with quasi-hyperbolic discounting, which predicts that violations of constant discounting only occur when the present (time point 0) is involved and, hence, not in sequence H2. We could also test quasi-hyperbolic discounting by removing the first observation (the present) from sequence H1. Then all health outcomes occur in the future and quasi-hyperbolic discounting predicts constant impatience. This prediction could also be rejected (Page's L-test, $p < 0.01$). Panels C and D show the time curves for the two money sequences M1 and M2. We could also reject constant impatience for money in favor of decreasing impatience (Page's L-test, both $p < 0.01$). The rejection of constant impatience in sequence M2 also violates quasi-hyperbolic discounting. Moreover, we could also reject the prediction of quasi-

[3]Decreasing impatience predicts that the WTW increases over the time trade-off sequence, which was largely confirmed. In the first health sequence (H1), the first and the second WTW were lower than the third and the fourth WTW (Wilcoxon test, all $p < 0.01$), but the first WTW did not differ from the second WTW and the third WTW did not differ from the fourth WTW. In the second health sequence (H2) all predictions were confirmed (Wilcoxon test, $p = 0.02$ in the comparison between the first and the second WTW, all other $p < 0.01$) except that the third and the fourth WTW did not differ.

hyperbolic discounting that constant impatience should hold in sequence M1 when the first observation is removed (Page's L-test, $p = 0.01$).

Figure 4.3: The elicited time trade-off sequences using the mean data

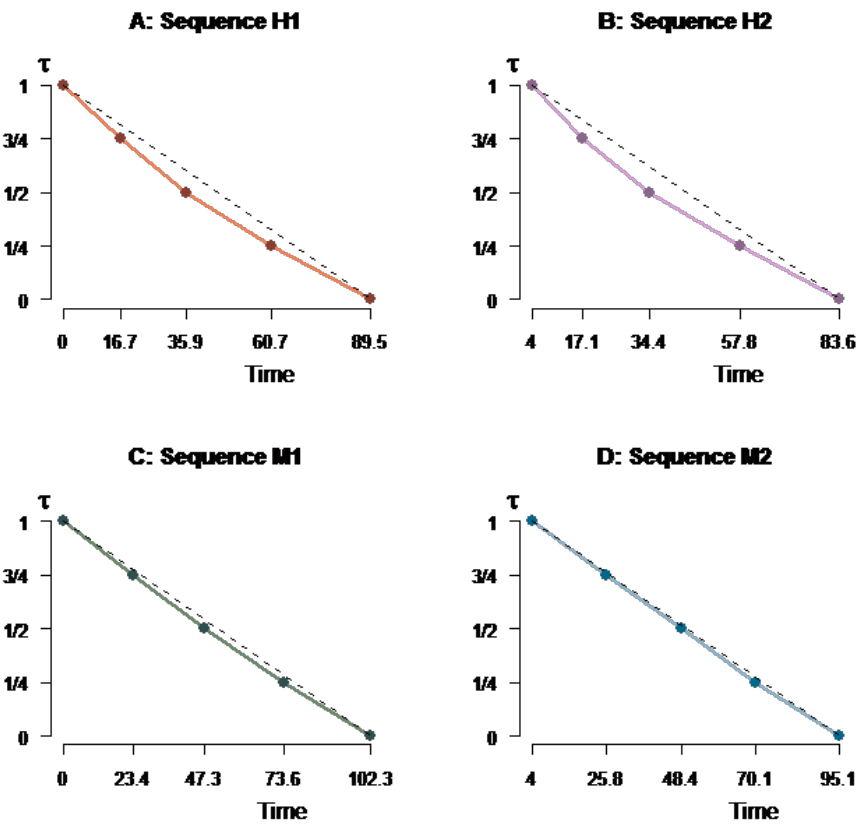


Figure 4.3 suggests that the deviations from constant discounting were larger for health than for money. To test this conjecture, we fitted the time curves by an exponential function $\tau(t) = e^{-\alpha t}$, where a reflects the convexity of the time curve and, thus, the degree of decreasing impatience $-\tau''/\tau' = -ln(\phi)''/ln(\phi)'$. We used R (Team, 2013) and the nlme package Pinheiro et al. (2007) to perform a nonlinear mixed-effects estimation of the exponential function with dummies to test for the fixed effects of outcome domain and the initial delay and subject as a random effect. P-values were obtained by likelihood ratio tests of the model with the dummy in question against the model without the dummy in question.

The exponential coefficients α were indeed larger for health than for money (LR-test, $p < 0.01$). We could not reject the null that the coefficients were the same for the two health sequences (LR-test, $p = 0.47$), but for money the coefficient of sequence M2 exceeded that of sequence M1 (LR-test, $p < 0.01$) signaling larger deviations from constant discounting in sequence M2. This finding, once again, contradicts quasi-hyperbolic discounting.

4.4.3 Individual results

The individual time curves showed much heterogeneity. To illustrate, Figure 4.4 shows the time curves of four subjects for sequence H1. Subject 24 was clearly decreasingly impatient and Subject 10 was clearly increasingly impatient. The time curve of Subject 26 resembles quasi-hyperbolic discounting. His willingness to wait first increases, which is consistent with decreasing impatience, and then remains constant. Finally, Subject 60 is first decreasingly impatient and then becomes increasingly impatient. To quantify decreasing impatience, we computed for each subject a *decreasing impatience (DI) index*. The DI index measures the difference between the area under the diagonal and the area under the normalized time curve [4]. It is positive for a decreasingly impatient subject, zero for a constantly impatient subject, and negative for an increasingly impatient subject. However, a subject with a zero DI index is not necessarily constantly impatient. Subject 60 is a case in point. His time curve crosses the diagonal and his DI index was close to zero, but he did not behave according to constant impatience. To capture such preferences, we also computed a nonstationarity (NS) index, which measures the deviation from constant impatience [5]. The NS index is related to the absolute value of the difference between the area under the diagonal and the area under the normalized time curve. The larger this area the more a subject deviated from constant discounting. The analysis based on the NS index mostly gave the same results and we only present the results when they differed from those based on the DI index.

[4] The normalized values of t_i are: $\tilde{t_i} = \frac{t_i - t_0}{t_4 - t_0}$, $i = 1,...,4$. They lie between 0 and 1. The normalization was necessary to compare the indices between subjects and to perform statistical tests. The values of t_i differed across subjects and, ceteris paribus, larger values of t_i lead to larger differences between the areas under the diagonal and under the time curve. The DI index is defined as: $DI = \sum_{i=1}^{4} (\frac{i}{4} - \tilde{t_i})$.

[5] The NS index is defined as: $NS = \sum_{i=1}^{4} |\frac{i}{4} - \tilde{t_i}|$.

Figure 4.4: Four different time curves

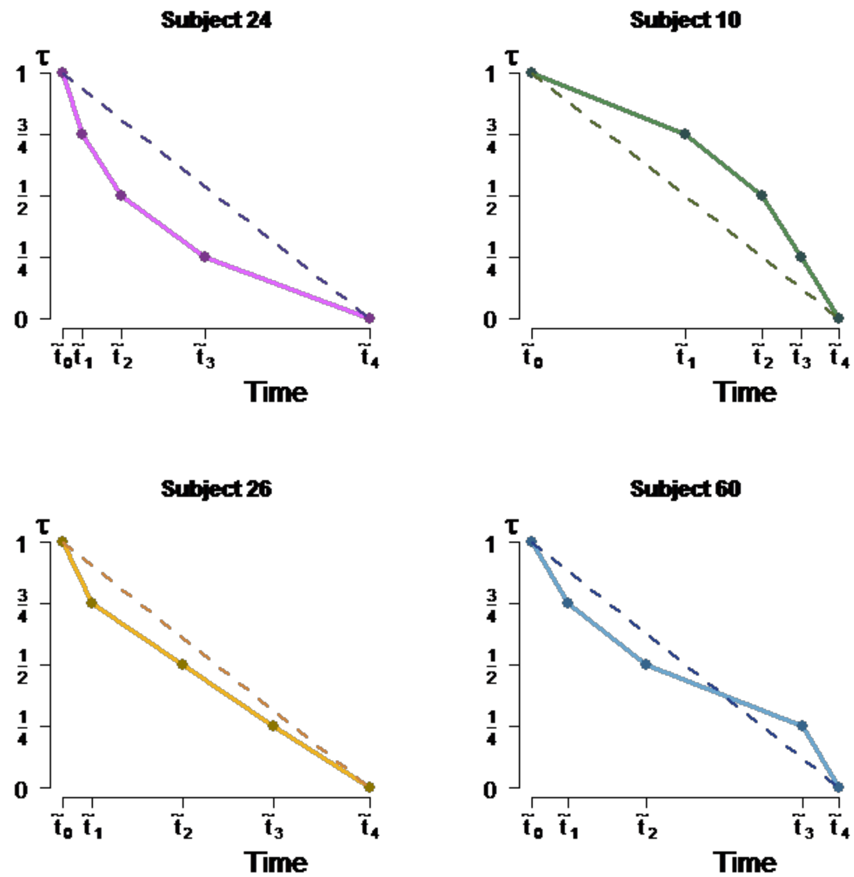


Table 4.3: Classification of subjects for health based on the DI indices

		Sequence H2			
		Decr.imp.	Incr.imp.	Const.imp.	*Subtotal*
Sequence H1	Decr.imp.	**27**	5	4	***36***
	Incr.imp.	4	**10**	2	***16***
	Const.imp.	1	1	**9**	***11***
	subtotal	***32***	***16***	***15***	**Total**: 63

Table 4.3 shows that in both health sequences a majority of the subjects were decreasingly impatient. Around 25% of the subjects (16 out of 63) were increasingly impatient. Impatience was a stable behavioral trait, as around 75% of the subjects displayed the same type of impatience in both sequences. Only 9 subjects switched from decreasing impatience in one sequence to increasing impatience in the other. These switching subjects were typically much more patient than non-switching subjects and,

as a result, their willingness to wait was somewhat imprecise. For example, subject 5's H1 and H2 sequences were (39,76,116,162) and (37,81,123,156). These sequences suggest that subject 5 was willing to wait around 40 weeks for the improvement in health, but as he was patient, he did not care much whether he had to wait, say, 38 weeks or 42 weeks. As a result, his responses were a bit imprecise and we classified him as decreasingly impatient in the first sequence and as increasingly impatient in the second sequence.

Table 4.4: Classification of subjects for money based on the DI indices

		Sequence M2			
		Decr.imp.	Incr.imp.	Const.imp.	*Subtotal*
Sequence M1	Decr.imp.	**22**	6	2	***30***
	Incr.imp.	10	**10**	2	***22***
	Const.imp.	1	1	**9**	***11***
	subtotal	***33***	***17***	***13***	**Total**:63

Table 4.4 shows that decreasing impatience was also the most common pattern for the two money sequences, but that between 25% and 35% of the subjects were increasingly impatient. Impatience was less stable for money than for health, with 16 subjects switches from decreasing impatience in one sequence to increasing impatience in the other. As for health, switching subjects were on average much more patient than non-switching subjects and the switch from decreasing to increasing impatience (and vice versa) could be explained by preference imprecision. However, there were also some subjects for whom the switch was probably caused by errors. For example, subject 56's M1 and M2 sequences were (4,8,12,15) and (7,10,14,17). These responses could be explained as follows. In sequence M1 he was always willing to wait 4 weeks, but he made one error such that his final willingness to wait was only 3 weeks. In sequence M2 he was always willing to wait 3 weeks, but, again, he made one error such that his third willingness to wait was 4 weeks. The consequence of these two small errors was that we classified him as increasingly impatient in sequence M1 and as decreasingly impatient in sequence M2 even though he, probably, was constantly impatient.

Figure 4.5: DI indices for the two health sequences

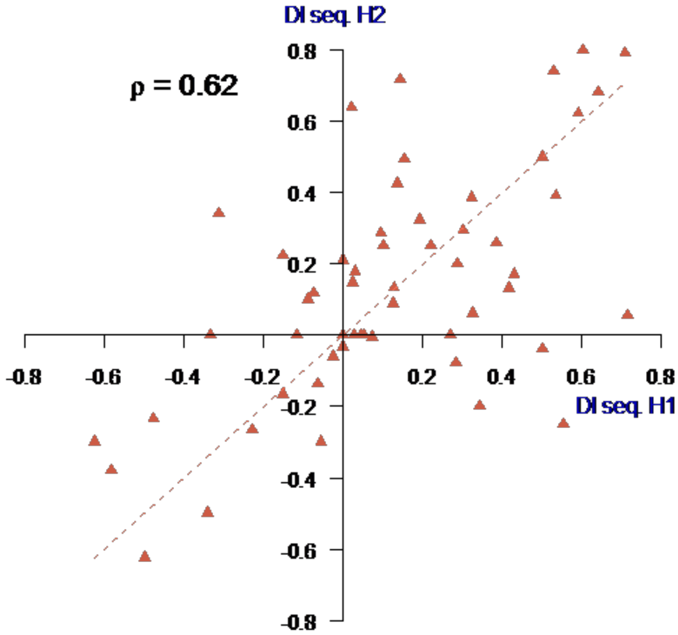


Figure 4.5 shows the DI indices for the two health sequences. Most points lie in the first quadrant and the mean DI index was significantly positive for both sequences (Wilcoxon test, both $p < 0.01$) consistent with decreasing impatience for health. The DI indices of sequences H1 and H2 did not differ (Wilcoxon test, $p = 0.54$). The correlation between the two DI indices was substantial ($\rho = 0.62$, $p < 0.01$), suggesting that time preferences for health were relatively stable.

Figure 4.6: DI indices for the two money sequences

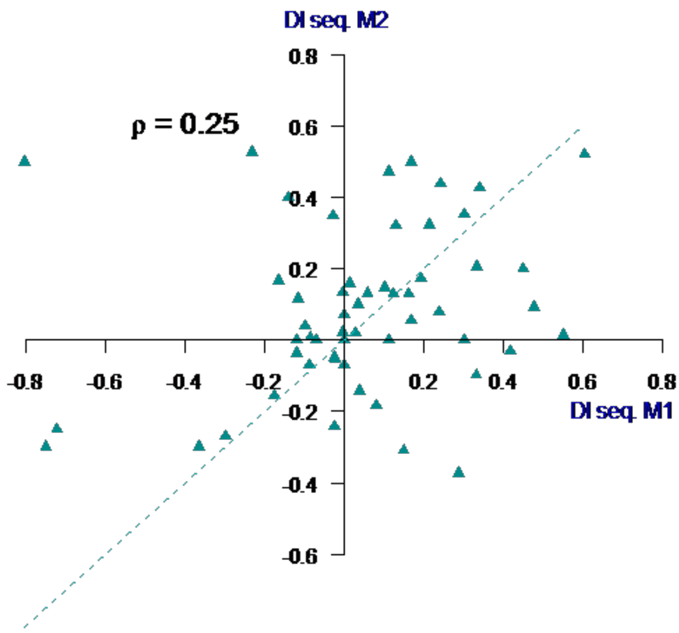
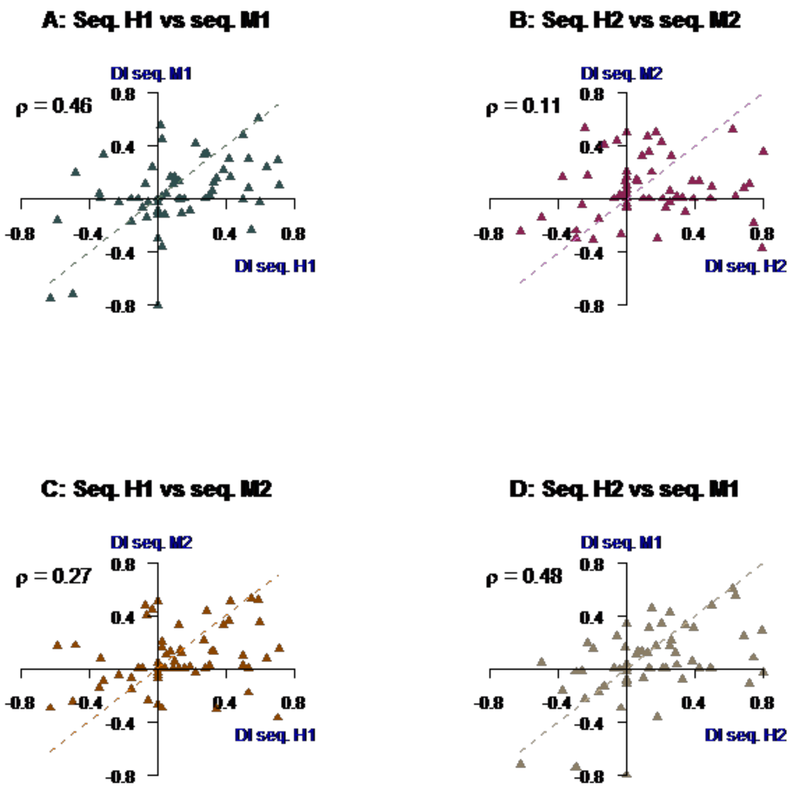


Figure 4.6 shows the DI indices for the two money sequences. The means of both sequences differed from 0 (Wilcoxon test, $p = 0.04$ for sequence 3 and $p < 0.01$ for sequence 4). The DI indices did not differ (Wilcoxon test, $p = 0.38$), but their correlation was only fair ($\rho = 0.25$,$p = 0.05$), suggesting that time preferences were less stable for money than for health. Figure 4.7 shows the relation between the DI indices for health and money. Panels A and D show that the indices for health and the first money sequence were moderately correlated ($\rho = 0.46$ ($p < 0.01$) between sequences H1 and M1; $\rho = 0.48$ ($p < 0.01$) between sequences H2 and M1). The correlations between the two health sequences and sequence M2 were much lower and only slight to fair ($\rho = 0.27$ ($p = 0.04$) between sequences H1 and M2; $\rho = 0.11$ ($p = 0.37$) between sequences H2 and M2).

Figure 4.7: DI indices for health versus money sequences



The DI indices indicated more decreasing impatience for health than for money. However, this difference was only significant in the comparison with sequence M1 [6]. The NS indices showed greater deviations from constant impatience for health than for money in all comparisons [7]. The finding that the DI and NS indices of sequences H2 and M2 differed significantly from zero provided further evidence against quasi-hyperbolic discounting.

[6] Wilcoxon tests, $p = 0.02$ in the comparison between sequences H1 and M1, $p = 0.18$ in the comparison between H1 and M2, $p = 0.03$ in the comparison between sequences H2 and M1, and $p = 0.19$ in the comparison between sequences H2 and M2.

[7] Wilcoxon tests, $p = 0.02$ in the comparison between sequences H1 and M1, $p = 0.01$ in the comparison between H1 and M2, $p = 0.06$ in the comparison between sequences H2 and M1, and $p = 0.03$ in the comparison between sequences H2 and M2.

4.4.4 Hyperbolic factors

So far, the analysis has shown that our subjects violated constant and quasi-hyperbolic discounting for both health and money. To gain additional insight into the validity of the different discount models, we computed hyperbolic factors for each subject. The hyperbolic factors are undefined when their denominator is negative, which happens when a subject is extremely decreasingly impatient. Such behavior cannot be accommodated by any of the hyperbolic alternatives for constant discounting and requires more general discount models (Ebert and Prelec 2007, Bleichrodt et al. 2009). It was rare in our data. For each subject, we computed 24 hyperbolic factors, 6 per sequence. In all sequences, less than 10% of the hyperbolic factors had a negative denominator (5% in H1 and M1, 8% in H2 and M2). We could not reject the null hypothesis of equal hyperbolic factors within each of the four sequences (Friedman test, all $p > 0.09$). However, this result was sensitive to the exclusion of the 6 most impatient subjects: without these subjects we could reject the null hypothesis of equal hyperbolic factors in sequence H2 (but not in the other three sequences). While most median hyperbolic factors equaled zero, all sequences contained at least one hyperbolic factor that was significantly different from 0 at the 1% significance level. This confirms, once again, that constant discounting and quasi-hyperbolic discounting did not hold. Finally, we could reject the prediction of power discounting that the hyperbolic factors equal 1 (Wilcoxon test, all $p < 0.01$). Consequently, the only models that were consistent with our data are Loewenstein and Prelec's (1992) hyperbolic discounting and Mazur's (1987) proportional discounting, except ,perhaps, for sequence H2.

4.5 Discussion

The novelty of this paper is that we quantify deviations from constant discounting across two domains: health and money. This quantification gives new insights into subjects' vulnerability to self-harming behavior and whether this vulnerability is domain-specific. Our tests make no assumptions about utility and do not require intertemporal separability. The main findings are as follows. First, our average subject deviated from constant discounting for both health and money and impatience decreased over time. However, time preferences were heterogeneous and a substantial minority of our subjects, between twenty-five and thirty-five percent, displayed increasing impatience. For money, our findings on increasing impatience confirm previous evidence (e.g., Abdellaoui et al., 2010; Sayman and Öncüler, 2009; Loewenstein, 1987; Takeuchi, 2011; Chesson and Viscusi, 2003; Rubinstein, 2003). For health, only indirect evidence of increasing impatience existed. The deviations from constant discounting were more pronounced for health than for money, which indicates that people might be more vulnerable to self-control problems for health than for money. This finding also suggests that intertemporal preferences are context-dependent and that findings for money outcomes cannot be simply transferred to health. We may have found less decreasing impatience for money due to market forces. As money is tradable on financial markets and transferable across time, people's discounting of money may have been disciplined by the prevailing interest rates for money (Cubitt and Read, 2007). However, we also found that time preferences were less stable for money than for health, which seems to contradict the above conjecture.

Another possible reason for the difference between discounting for health and money may be that our subjects were more familiar with decisions about money than decisions about health. Yet, one would then, again, expect that time preferences would be more stable for money than for health, in contrast to our observation. Moreover, Chapman et al. (1999) found that familiarity with medical treatments did not increase the similarity between time preferences for money and health.

Possibly, time preferences were more stable for health than for money because subjects perceived their future health as more stable than their future income. For delays up to 10 years, which we considered in our experiment, health is usually rather stable and, moreover, we instructed subjects to assume that their baseline health would remain constant to chronic back pain. However, we did not tell subjects that their income would remain constant and, as they were students, substantial changes in their income were likely in the near future (after graduation). Consequently, subjects may have realized that their future evaluation of money would change. Our analysis assumed that the evaluation of money would not change over time and this may have affected the estimations. For a more detailed analysis of the impact of time-dependence of utility on discounting behavior see Gerber and Rohde (2010, 2015).

Our final contribution is to obtain new evidence on the descriptive validity of discount models. Such evidence is still scarce, particularly for health. Hyperbolic discounting (Loewenstein and Prelec 1992) and proportional discounting best described intertemporal preferences, but it should be kept in mind that these two models cannot accommodate the behavior of the increasingly impatient subjects. To explain increasing impatience other discount functions are needed (Ebert and Prelec, 2007; Bleichrodt et al., 2009). The widely-used quasi hyperbolic discounting model was rejected for both money and health, casting doubt on the descriptive realism of studies that use this model to derive predictions about behavior.

Chapter 5

Cash in Hand, Crashes in Mind: Cash Aggravates Probability Weighting

5.1 Introduction

Nowadays, consumers are indulged with many payment instruments: cash, checks, debit cards, credit cards, online/mobile banking, Apple Pay, etc. Cash is no longer the most common payment instrument in Europe or in the U.S. (Bagnall et al., 2014). Non-cash payment instruments enable consumers to make payments without exchange of cash notes, which substantially simplify the payment process. Non-cash payment instruments obviously change how we pay. What is not as obvious but no less important is, they also change how much we pay.

When payments are made in cash, consumers tend to spend less (Hirschman, 1979; Feinberg, 1986). Even for hypothetical questions, consumers cued with credit card logo in sight are willing to spend more and have shorter decision time (Feinberg, 1986; Raghubir and Srivastava, 2008).The substantial gap between cash and non-cash payment cannot be fully explained by the convenience and potential money saving of credit-card usage. A behavioral explanation is that the thoughts of payment can undermine the pleasures of consumption, and the psychological distance created by non-cash

This chapter is based on the homonymous paper, co-authored with Ning Liu.

payment could alleviate the pain of paying (Prelec and Loewenstein, 1998).

To test the effects in incentivized transactions of high value, Prelec and Simester (2001) conducted experiments comparing consumers' willingness-to-pay for tickets to sporting events with different payment instruments. Consistent with the literature, consumers were willing to pay substantially more with credit card. The large credit card premium (up to 100%) was implausible to be explained by liquidity constraints. However, the effect seemed to depend on the characteristics of the products under valuation. In their second study, subjects were asked to value products of either certain (a restaurant gift certificate) or unknown market value (tickets to a sold-out sporting event, for which the value was unstated). The gap between cash and non-cash instruments only existed for the products with unknown market value, but not for those with certain market value.

If the effects of payment instruments are moderated by the feeling of uncertainty, the question arises as to through which channel do the payment instruments work. Prelec and Simester (2001) did not provide a theory explaining the presence of the observed effect. In this chapter, we propose that payment instruments change valuations of lotteries through shaping consumers' risk attitudes. In particular, consumers' probability weighting might be affected by payment instruments in two ways. On the one hand, payment instruments could affect consumers' attention allocation. As Kahneman (2011) put it, "our mind has a useful capability to focus on whatever is odd, different or unusual". The attention paid to the colorful cash occupies cognitive resources. Different notes and coins make it one-step harder to calculate EV of the lottery. The depletion of cognitive resource reduces peopleąŕs reliance on the analytic, calculating, and deliberative so called "System 2", and rely more on the instinctive "System 1". Therefore people would be less sensitive to probability differences. On the other hand, risk-as-feelings hypothesis (Loewenstein et al., 2001) postulates that responses to risky situations result in part from direct emotional influences, including (negative) feelings such as worry, fear, or anxiety, and such feelings can be influenced by how an outcome is presented (vividness). Compared with non-cash presented lotteries, cash presented ones might trigger stronger anticipatory emotions and therefore make people more pessimistic towards risk.

We test the hypotheses above using a controlled laboratory experiment. We ask the subjects to value lotteries with known probabilities under two treatments, one with cash and the other with a non-cash instrument (number). We found that the valuations given by subjects in the cash treatment were lower than those in the number treatment. We use a binary rank-dependent utility (RDU) model to explain the certainty equivalents (CEs) given under the two treatments. Since we used binary lotteries in the experiment, many non-expected utility theories do not diverge (Gul, 1991; Luce and Fishburn, 1991; Miyamoto, 1988), and therefore the results from binary RDU also apply to them. By eliciting the parameters of the utility function and probability weighting function under each treatment, we identified that the gap in CEs between the treatments was due to the difference in probability weighting functions under the cash and non-cash treatments. Subjects were less sensitive to changes in likelihood when valuing cash lotteries, however there is no difference in pessimism.

5.2 Experimental design

Ninety-two students at Erasmus University Rotterdam participated in the experiment (37% female). Each subject received a show-up fee of €5. On top of that, each subject received additional payment (up to €30) determined by their choice in a randomly drawn question answered in the experiment.

Subjects were assigned to one of the two treatments randomly, and were interviewed individually by one of the two experimenters randomly determined, independent of the treatment. In each session, the experimenter presented to the subject a series of lotteries, and recorded their valuation to each. To familiarize subjects with the tasks and payment procedure, the instructions contained examples and trial problems. The subjects could ask the experimenter clarification questions any time during the experiment. To minimize the difference between the two experimenters, a strict protocol (see Appendix) about what to tell subjects and how to answer their questions was adopted. Subjects could work at their own speed. On average, it took them 45 minutes to complete the experiment.

In both treatments, subjects were asked to give valuations to binary lotteries. We denote $L = x_p y$ (with $x > y > 0$). The lottery gave the subject the better outcome x with probability p, and y otherwise. There were in total 12 such lotteries, varying p, x and y (see Table 5.1). Such variation enables us to estimate the utility function and the probability weighting function for each subject. The lotteries appeared in individualized random orders.

Table 5.1: The lotteries used in the valuation task

	p	x	y
1	0.05	20	5
2	0.05	30	10
3	0.1	10	5
4	0.25	20	5
5	0.25	30	10
6	0.5	10	5
7	0.5	20	5
8	0.75	20	5
9	0.75	30	10
10	0.9	10	5
11	0.95	20	5
12	0.95	30	10

We implemented the Becker-DeGroot-Marschak (BDM) procedure (Becker et al., 1964) to elicit CEs for lotteries with compatible incentives. First, the lottery to be implemented for real was randomly determined at the end of the experiment. Second, the BDM procedure was conducted by drawing one number z between the lowest prize (y) and the highest prize (x) of the chosen lottery. If z was larger than the subject's evaluation, the subject would receive z; otherwise, the subject would be paid by running the lottery.

Figure 5.1 shows how the lotteries were presented to subjects in the two treatments. The only difference between the two treatments is that in the number treatment, money amounts were presented with the currency symbol and a number, as in most of the experiments in decision studies; in the cash treatment, money amounts were presented with real cash attached to the questionnaire without the number written down[1].

[1] In the cash treatment, the same money amount could be presented in different ways. For instance, €13 could be presented with two €5 notes plus three €1 coins, or with six €2 coins plus one €1 coin, or other possible ways. We apply the rule that in the cash treatment, money amount is presented with the fewest number of notes and coins. In this case, €13 is presented with one €10 note, one €2 coin and one €1 coin.

Figure 5.1: Examples of lottery presentation

(a) Number Treatment

(b) Cash Treatment

Subjects were asked to either specify their evaluation to the lottery by writing down the number in the number treatment, or put down the corresponding amount of cash below the lottery in the cash treatment. Particularly, subjects in the cash treatment were given a box with one €20 note, one €10 note, one €5 note, two €2 coins, one €1 coin, one 50-cent coin, two 20-cent coins and one 10-cent coin in it, so that they can make different combinations to present all possible evaluations (precision up to 10 cents) to lotteries, ranging from €41 to 0.

5.3 Analysis

5.3.1 Decision-model-free analyses

We analyze the reported CEs without assuming any specific decision model. We use a simple linear mixed-effects model, with fixed effects of treatments and task dummies and with subject random effect. The certainty equivalent of lottery j given by subject i is modeled as: $CE_{ij} = \beta\text{Treatment}_i + \delta_j + \epsilon_{ij}$, where $i = 1, ..., 92, j = 1, ..., 12$, and ϵ_{ij} is a normally distributed within-subject error term.

We also calculate Relative Risk Premium ($RRP = \frac{EV-CE}{EV}$) for each valuation. The positive, zero, and negative RRPs suggest risk aversion, risk neutrality and risk seeking respectively. We model the RRPs with the same mixed effect model as the one for CE above: $RRP_{ij} = \beta\text{Treatment}_i + \delta_j + \epsilon_{ij}$, where $i = 1, , 92, j = 1, , 12$, and ϵ_{ij} is a normally distributed within-subject error term.

5.3.2 Binary RDU analysis

Under binary RDU, for a given binary lottery $L = x_p y$ $(x > y \geq 0)$, the CE shall satisfy: $CE = u^{-1}(w(p)u(x) + (1 - w(p))u(y))$ (Eq.5.1), where u is a utility function, with $u(0) = 0$ and $u'(x) > 0$, describing how a monetary outcome x is subjectively valued, and w is an increasing probability weighting function that assigns subjective weight to probabilities, with $w(0) = 0$ and $w(1) = 1$.

Stott (2006) compared combinations of different utility functions and weighting functions for choice data, and found that the combination of power utility function (Wakker, 2008) and the compound invariance family (Prelec, 1998) the most predictive. We therefore use the power utility function $u(x) = x^\gamma$ if $\gamma > 0$; $\ln x$ if $\gamma = 0$; $-x^\gamma$ if $\gamma < 0$ (Eq.5.2), and Prelec's compound invariant probability weighting function $w(p) = ((exp(-(-\ln p)^a))^b$ $(0 < a < 1,\ b > 0)$ (Eq.5.3) to analyze our data. In particular, we use Prelec's two parameter probability weighting function that decomposes probability weighting into likelihood-sensitivity and pessimism.

The parameter a is an index of likelihood-sensitivity, which points to a psychological phenomenon which reflects "diminishing sensitivity" for probabilities. A smaller a

indicates less distinction between different levels of likelihood. The parameter b is an index of pessimism, and a bigger b indicates that the subject pays more attention to the worst outcome.

Using maximum-likelihood estimation, we estimate Eq. 5.1 with the specific u and w for each individual separately, and obtain parameters γ, a, and b for each individual. We will compare the estimates to their benchmark and between the two treatments using Wilcoxon tests.

5.4 Results

5.4.1 Decision-model-free analyses

Table 5.2 shows the means (and standard deviations) of CEs for each lottery in the two treatments. The EVs, winning probability of the larger outcome, the mean differences between treatments normalized by EVs of the lotteries are also provided in the table.

Table 5.2: CEs for each lottery

	EV	$Probability$	$Number$	$Cash$	$\frac{Difference}{EV}$
1	5.5	0.1	6.20 (0.94)	6.05 (0.82)	2.72%
2	5.75	0.05	7.60 (2.66)	7.50 (2.34)	1%
3	7.5	0.5	7.60 (0.86)	7.39 (0.55)	3.87%
4	8.75	0.25	9.58 (2.75)	9.17 (2.33)	4.69%
5	9.5	0.9	8.55 (0.99)	8.40 (1.09)	1.58%
6	11	0.05	13.64 (3.01)	12.61 (2.26)	9.36%
7	12.5	0.5	12.50 (2.42)	12.30 (1.60)	1.60%
8	15	0.25	15.98 (2.76)	15.22 (2.71)	5.07%
9	16.25	0.75	15.50 (1.95)	14.65 (2.59)	5.23%
10	19.25	0.95	17.71 (1.58)	17.34 (2.31)	1.92%
11	25	0.75	23.10 (2.43)	20.46 (4.21)	10.56%
12	29	0.95	26.97 (2.25)	24.47 (4.94)	8.62%

It can be observed that CEs in the number treatment are larger than CEs in the cash treatment for every lottery. As described in Section 5.3.1, we subject the CEs to a linear mixed-effects model. The model shows that the CEs in the cash treatment are on average 0.77 euro lower than those in the number treatment ($p = 0.002$).

If look at the columns of EV and $\frac{Difference}{EV}$, an increasing trend can be detected: the normalized difference between two treatments is increasing with EV. This trend is shown with a fitted line in Figure 5.2. Pearson correlation test also confirms this trend ($\rho = 0.62, p - value = 0.03$).

Figure 5.2: Scatter plot of proportional difference sorted by EVs.

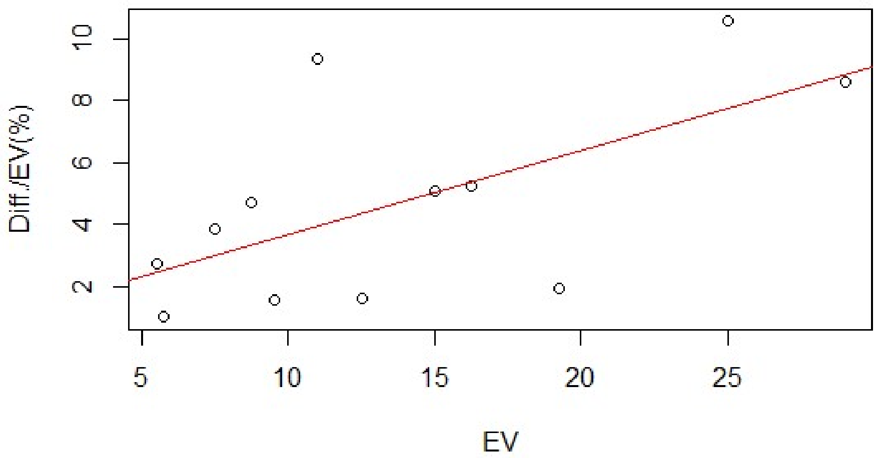


Figure 5.3 shows mean relative risk premium for lotteries with different probabilities. The mean $RRPs$, sorted by the probability p of outcome x_1, show a systematic relationship between risk attitudes and probabilities of outcomes, which is also consistent with the typical empirical findings: On average, people are risk seeking for small probabilities, and risk averse for large probabilities.

We subject the $RRPs$ to a linear mixed effects model, as described in Section 5.3.1. The model shows that the cash treatment increases the RRP by 0.05 ($p = 0.03$).

5.4.2 Binary RDU analysis

Table 5.3 summarizes the results from the maximum likelihood estimation specified in Section 5.3.2.

The results above show that, at the aggregate level, subjects in the number treatment exhibit linear utility, likelihood insensitivity and no pessimism. The cash treat-

Figure 5.3: RRP by the probabilities of the better outcome.

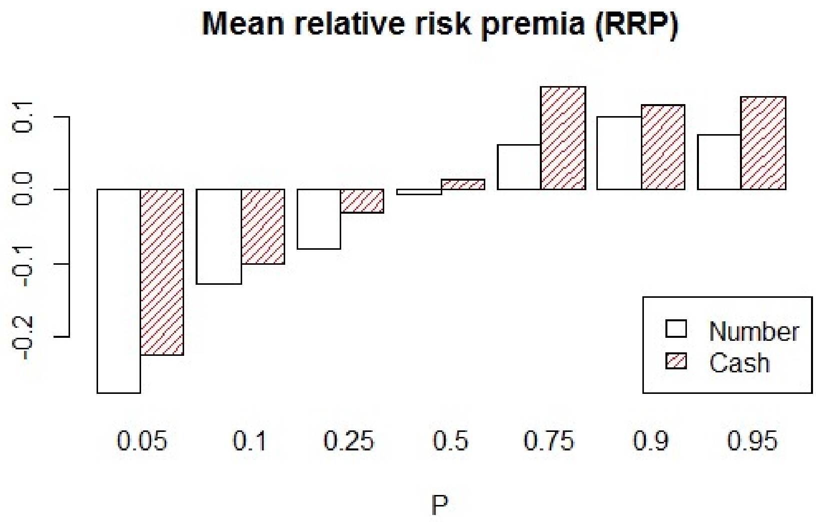


Table 5.3: Parameters from maximum likelihood estimation

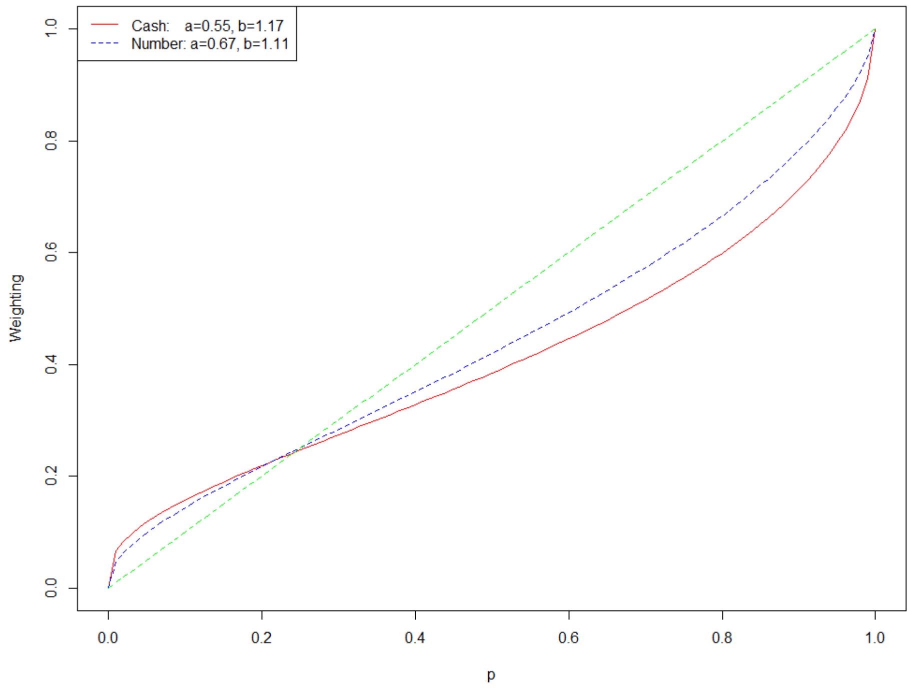
		Cash treatment	Number treatment	p-value
Utility Curvature γ	median	1.09	1.07	0.97
	mean	1.22	1.35	
Likelihood Sensitivity a	median	0.49***	0.69***	0.02
	mean	0.55	0.67	
Pessimism b	median	1.00	1.01	0.46
	mean	1.17	1.11	
Number of Observations		46	46	

Notes: Reported numbers are the medians and means of estimated coefficients in the corresponding treatment, followed by significance from one-sample Wilcoxon signed rank test. The benchmarks for the coefficients (γ, a, b) are 1.
The last column gives p-values from Wilcoxon rank-sum test.
*** indicates significance at 1% (two-sided test)

ment does not change the utility curvature, but changes the probability weighting. In particular, subjects in the cash treatment are less sensitive to probability changes than those in the number treatment.

In Figure 5.4, we plot the probability weighting curves (based on means of individual parameters from maximum likelihood estimation) for the two treatments. The curve of the cash treatment is more pronounced in its inverse S shape.

Figure 5.4: Probability weighting curves of the two treatments.



5.5 Conclusion

It has been noticed in the literature that the gap between valuations made of cash and non-cash instruments is more prominent for products of unknown value than for those of clear market value. Using a simple experiment, we test how payment instruments influence valuation through affecting people's risk attitudes, which can be reflected by utility curvature and probability weighting.

The results show that valuators' utility functions elicited for cash and non-cash payment instruments do not differ from each other. The difference in valuations is driven by probability weighting. Presenting lotteries with cash makes valuators less sensitive to changes in likelihood, which leads to less variation in valuations of different lotteries.

5.6 Appendix. Experimenter's protocol

1. Is this number how much I want to pay for / sell for this lottery?

   "This is a valuation task, and you are asked to fill in how much is this lottery worth to you. Our payment procedure is designed to guarantee that it is for your best interest to fill in the exact valuation in your mind, which dominates both overstating and understating this value." (Specifically avoid mentioning "buy" or "sell" in the explanation.)

2. In case the subject gives a valuation lower than the lower outcome in the lottery.

   "Sorry to interrupt. You can surely put whatever amount you see proper as your valuation. This is just a reminder, because here you put a valuation lower than the lower possible outcome in the lottery, and I want to clarify the rules in case there is any misunderstanding. Since we will only randomly draw a number from the lower outcome and the higher outcome of a given lottery, in this case X and Y (X<Y are the two outcomes of the lottery this subject is valuating), therefore giving a valuation lower than the lower outcome of the lottery means that all the random number we draw would be higher than your valuation and therefore you will be paid that amount. In the extreme case, if we draw X, your valuation indicates that you prefer to be paid X, rather than receiving this lottery that gives you at least X. Is this what you prefer?"

3. In case the subject gives a valuation higher than the higher outcome in the lottery.

   "Sorry to interrupt. You can surely put whatever amount you see proper as your valuation. This is just a reminder, because here you put a valuation higher than the higher possible outcome in the lottery, and I want to clarify the rules in case there is any misunderstanding. Since we will only randomly draw a number from the lower outcome and the higher outcome of a given lottery, in this case X and Y (X<Y are the two outcomes of the lottery this subject is valuating), therefore giving a valuation higher than the higher outcome of the lottery means that all the random number we draw would be lower than your valuation and therefore

you will receive the lottery. In the extreme case, if we draw Y, your valuation indicates that rather than receiving Y, you prefer to receive the lottery that gives you at most Y. Is this what you prefer?"

4. In the cash treatment, make sure the subject put all the notes and coins for valuation back to the box after finishing each valuation.

Chapter 6

Are Black Swans Really Ignored? Re-examining Decisions from Experience

6.1 Introduction

Studies of decisions from experience (henceforth, DFE) investigate decision situations in which people rely on personal experiences when facing uncertainty. Decision makers often have no access to possible choice outcomes, let alone to the corresponding probabilities. Instead, they make decisions based on the past observations in their memory. DFE better captures real life decisions than traditional "Decisions from Description" (henceforth, DFD) where payoffs and probabilities are fully specified, which rarely happens the case in real life. In the usual sampling paradigm of DFE (Hertwig et al., 2004), subjects learn about unknown payoff distributions by drawing samples with replacement. With merely these cases in memory, they make their final decisions.

Since Hertwig et al. (2004), an intriguing discrepancy between the two decision paradigms, which is called the DFE-DFD gap, has received plenty of attention. The common view in the DFE literature is that rare and extreme events, so called "black

This chapter is based on the homonymous paper, co-authored with Ilke Aydogan.

swans", are underweighted under the DFE paradigm whereas they are overweighted under the DFD paradigm (for a review, see Hertwig and Erev (2009)). This implies a complete reversal of the inverse S-shaped probability weighting that has been documented by many empirical studies under DFD (Abdellaoui, 2000; Bleichrodt and Pinto, 2000; Bruhin et al., 2010; Booij et al., 2010; Fehr-Duda et al., 2006; Gonzalez and Wu, 1999; Tversky and Kahneman, 1992; Wu and Gonzalez, 1996).

The DFE literature has indicated that the DFE-DFD gap is a robust empirical phenomenon. Although the under-sampling of rare events due to reliance on small samples mostly explains the early findings on the gap (Hadar and Fox, 2009; Fox and Hadar, 2006; Hertwig et al., 2004), later studies have shown that it does not provide a complete account (Barron and Ursino, 2013; Camilleri and Newell, 2009; Hau et al., 2010, 2008; Ungemach et al., 2009). Moreover, different attitudes towards risk (known probabilities in DFD) and ambiguity (partially unknown probabilities in DFE) are another cause of the gap (Abdellaoui et al., 2011; Kemel and Travers, 2015). Despite the robustness of the DFE-DFD gap, whether it can actually amount to a reversal - or only an attenuation - of the inverse S-shaped probability weighting is still unclear in the literature.

In addition to the sampling error and ambiguity, there are two extra confounds that render the inferences about probability weighting problematic in DFE studies. The first concerns an aggregation problem when there is a lack of control over the sampling experience of subjects. Because of the random nature of the sampling process - where the sampling is made with replacement and subjects decide when to stop sampling - each subject relies on her own distinct subjective experiences while making her choices in the sampling paradigm. Importantly, this heterogeneity at the individual level causes potential distortions at the aggregate level due to averaging artifacts (see Estes,1956; Estes,2002; Sidman,1952). We elaborate on this issue in the section of the DFE-DFD Gap.

The second confound concerns is regarding the role of utilities in the investigation of probability weighting. In proceeding studies of DFE, the underweighting of rare outcomes is typically inferred from the preference of sure gains over EV-equivalent lotteries with rare probability (for example, a preference of $1 for sure over a lottery with 10% chance of winning \$10 and \$0 otherwise). It seems that the prevalent risk seeking for unlikely gains under DFD turns into an aversion under DFE (see also the review of the DFE literature by Rakow and Newell 2010). However, it is important to recognize that the aversion to unlikely gains may as well be due to concave utility (possibly coupled with an unbiased probability weighting) as it may be due to an underweighting of unlikely events.

This paper provides a reliable measurement of probability weighting under DFE by resolving the aforementioned problems. First, we used Barron and Ursino's (2013) adjustment of the sampling paradigm to obtain a control over the sampling experience of each individual subject. Specifically, all of our subjects were required to carry out complete sampling from finite outcome distributions without replacement. Hence, they acquired the precise sampling information that matched with the objective probabilities without any sampling error or ambiguity[1].

Next, rather than relying on indirect inferences, we measured probability weighting by a rigorous two-stage methodology (Abdellaoui, 2000; Bleichrodt and Pinto, 2000; Etchart-Vincent, 2004; Qiu and Steiger, 2010). In particular, controlling for the utility curvature in the first stage, each choice in the second stage exactly revealed overweighting or underweighting of probabilities. Thus, our experimental setup enabled us to identify the direction and the magnitude of the deviations from expected utility (henceforth, EU), and hence find out what the exact DFE-DFD gap is.

6.2 Deviations from EU due to probability weighting

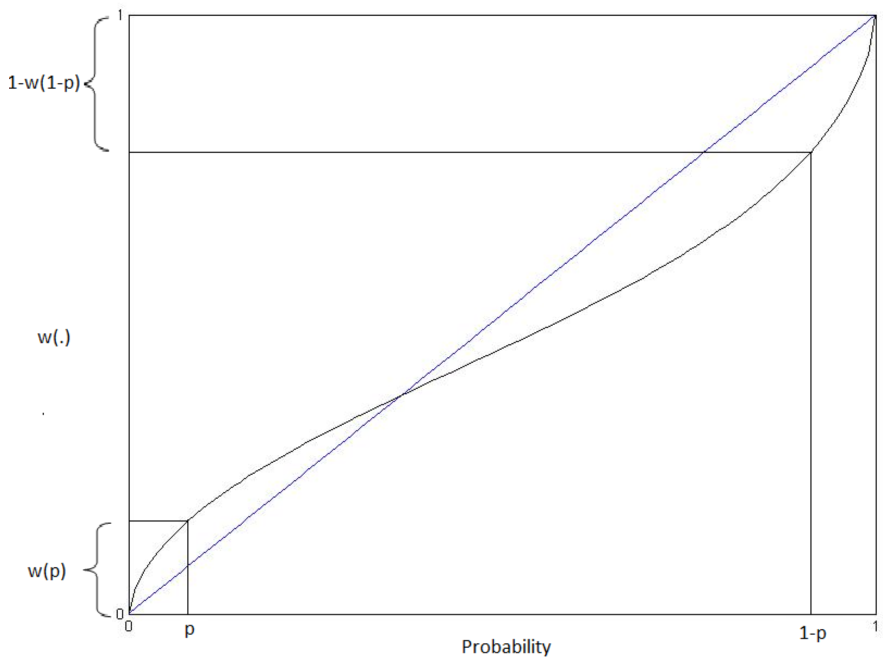
We restrict our attention to probability-contingent binary prospects in the gain domain. A binary prospect of winning α with probability p and β otherwise is denoted $\alpha_p\beta$. Under rank dependent utility (henceforth RDU), for $\alpha \succcurlyeq \beta \succcurlyeq 0$, $\alpha_p\beta$ is evaluated by $w(p)U(\alpha) + (1 - w(p))U(\beta)$ where U is the utility function and w the probability weighting function. Throughout, we assume binary RDU. Most other non-EU theories, in particular both versions of Prospect Theory for gains (Kahneman and Tversky, 1979;

[1]Barron and Ursino (2013) also investigates the DFE-DFD gap using the same adjustment of the sampling paradigm. A comparison of the current study with Barron and Ursino is provided in section 6.3.

Tversky and Kahneman 1992), and Gul's (1991) Disappointment Aversion Theory, agree with the binary RDU in the evaluation of binary prospects (Observation 7.11.1. in Wakker (2010), pp. 231). Hence, our analysis applies to all these theories.

RDU deviates from EU when $w(.)$ is not the identity. Thus, the risk attitude of a decision maker depend not only on the utility curvature as in EU but also on probability weighting. The common finding with the DFD paradigm is an inverse S-shaped (first concave and overweighting, then convex and underweighting) probability weighting function (Figure 6.1)[2]. The steepness of the probability weighting function at the both end points implies that the rare and extreme outcomes in general receive too much decision weight. When a rare outcome with probability p is desirable, its impact given by $w(p)$ is overweighted because of the overweighting of small probabilities ($w(p) > p$). This increases the attractiveness of the prospect, and enhances risk seeking. Similarly, when a rare outcome with probability p is unfavorable, its impact, given by $1-w(1-p)$, is overweighted because of the underweighting of large probabilities ($w(1-p) < 1-p$). This decreases the attractiveness of the prospect, and enhances risk aversion.

Figure 6.1: Inverse S-shaped probability weighting function



[2]For evidence against inverst-S, see Qiu and Steiger (2011), van de Kuilen and Wakker (2011) and Krawczyk (2015).

The pattern of inverse S-shaped probability weighting is commonly interpreted as the reflection of both cognitive and motivational deviations from EU (Gonzalez and Wu, 1999). On the one hand, the simultaneous overweighting and underweighting of extreme probabilities implies insufficient sensitivity to intermediate probabilities. This effect is called likelihood insensitivity, and points to cognitive limitations in discriminating different levels of uncertainty. On the other hand, underweighting of moderate probabilities (such as, $w(0.5) < 0.5$) suggests a pessimistic attitude towards risk in the major part of the probability domain. This effect points to motivational deviations from EU.

6.3 The DFE-DFD gap

Hertwig and Erev (2009) considers three DFE paradigms: partial feedback, full feedback, and sampling paradigms. The essential feature shared by all three DFE paradigms is that subjects learn about unknown payoff structures by solely relying on their past experiences. In the partial feedback paradigm, subjects learn about outcomes and probabilities by making repeated choices, and receiving feedback about the realized outcomes (Barron and Erev, 2003). In the full feedback paradigm, subjects also learn about the forgone outcomes from the unchosen options (Yechiam and Busemeyer, 2006). The sampling paradigm involves a single, rather than repeated, choice preceded by a purely exploratory and inconsequential sampling period in which subjects draw outcomes from unknown payoff distributions with replacement, usually as many times as they wish (Hertwig et al. 2004; Weber et al. 2004).

All three paradigms lead to similar behavioral patterns with an apparent underweighting of rare and extreme outcomes, which contradicts the common empirical findings from DFD. However, although the empirical findings with all three paradigms are alike, the two feedback paradigms are inherently different from the sampling paradigm (for an empirical comparison of three DFE paradigms, see Camilleri and Newell 2011, but also see the theoretical discussion of Gonzalez and Dutt 2011). In particular, repeated choices in the two feedback paradigms, as opposed to single decisions in the sampling paradigm, induce long run pay-off considerations due to accumulating income. This is expected to lead to higher rates of expected value maximization in repeated choices by the law of large numbers (Keren and Wagenaar, 1987; Lopes, 1982; Tversky and Bar-Hillel, 1983). Furthermore, distinct psychological factors, such as reinforcement learning, and the hot stove effect (production of success), also play a role in repeated decisions with feedback (March, 1996; Denrell and March, 2001). Erev and Barron (2005) also reviews the effects which lead to deviations from expected value maximization in repeated choice paradigms. The sampling paradigm, on the other hand, is more comparable with the DFD paradigm as both involve single decisions. Therefore, the intriguing gap between the sampling paradigm and DFD has received most attention in the DFE literature. The current paper also focuses on the sampling paradigm of DFE.

6.3.1 The information asymmetry account and the sampling error

The main premise of the DFE-DFD gap is in which the way that the information about uncertain prospects is acquired matters in decisions under uncertainty. In other words, experience matters (Hau et al., 2008).

Fox and Hadar (2006) and Hadar and Fox (2009) argue that there is an important caveat associated with this premise. DFE and DFD differ from each other not only in terms of the way that the information is acquired but also in terms of the information available to subjects. Indeed, whereas the precise probabilities and outcomes are known in DFD, they remain partially unknown in DFE. This means that subjects in DFE have to rely on their own subjective probability judgments based on the sampling information they acquire. Importantly, subjective probabilities are prone to diverge from objective probabilities due to potential errors either in the sampling process or in subjective probability judgments. This generates an information asymmetry between DFE and DFD. Hadar and Fox (2006) indicates that the underweighting of rare outcomes observed by Hertwig et al. (2004) is almost entirely caused by the sampling error as subjects often under-observe, or even never observe, the rare outcomes due to reliance on small samples. On the other hand, judgment error and underestimation of rare outcomes are not found to be a significant sources of the gap.

Later studies test this information asymmetry account of the DFE-DFD gap by reducing or completely eliminating the sampling error. Several papers demonstrated that the gap is actually persistent when the subjects are obliged to draw large or even representative samples from underlying probability distributions (Barron and Erev, 2003; Camilleri and Newell, 2009; Hau et al., 2010, 2008; Ungemach et al., 2009). Moreover, subjective probability judgments are usually found well calibrated although their correlation with observed relative frequencies is imperfect (Camilleri and Newell, 2009; Hau et al., 2008; Ungemach et al., 2009; Barron and Yechiam, 2009). These findings suggest that the DFE-DFD gap is not just an artifact of information asymmetries between the two cases but indeed a robust psychological phenomenon.

6.3.2 DFE and DFD: two different sources of uncertainty

Nevertheless, although the aforementioned studies solve the problem of sampling error, the uncertainty about the outcome probabilities remains. This residual uncertainty about probabilities makes DFE a case of ambiguity whereas DFD is a case of risk. Abdellaoui et al. (2011) and Kemel and Travers (2015) shows that ambiguity attitudes, i.e. the different attitudes towards known vs. unknown probabilities, also play a role in the DFE-DFD gap. These studies investigate the role of ambiguity by an intermediate design between DFE and DFD. In particular, while adhering to free sampling, they also provide subjects with the list of outcomes that the prospects involved to avoid the cases of ignorance[3] (also see the incomplete information treatment of Kemel and Travers, 2016). Both studies replicate the well-known gap but do not find evidence against inverse S-shaped probability weighting under DFE. The absence of underweighting in these studies may be explained by the mere presentation of outcomes in their modified DFE design, which increases the salience of the rare outcomes (Erev et al. 2008, see also the discussions in Abdellaoui et al. 2011 and Kemel and Travers 2015).

[3]Ignorance refers to the lack of knowledge about possible outcomes here.

6.3.3 Problem of aggregation in the sampling paradigm

Besides the sampling error and the uncertainty about outcome probabilities, there is also a methodological difficulty in making inferences about the impact of experience in the sampling paradigm. This difficulty concerns the aggregation of individual choices when there is substantial heterogeneity in sampling experiences of subjects. As explained above, experienced probabilities differ from objective probabilities either due to sampling error or due to judgment errors. As a result, each subject makes her choice based on her own subjectively experienced probabilities. Notably, as the aggregation of such individual choices amounts to taking the average of the weightings, rather than the weighting of the average, of experienced probabilities, the concave-convex curvature of the inverse S-shaped probability weighting function may lead to an erroneous DFE-DFD gap.

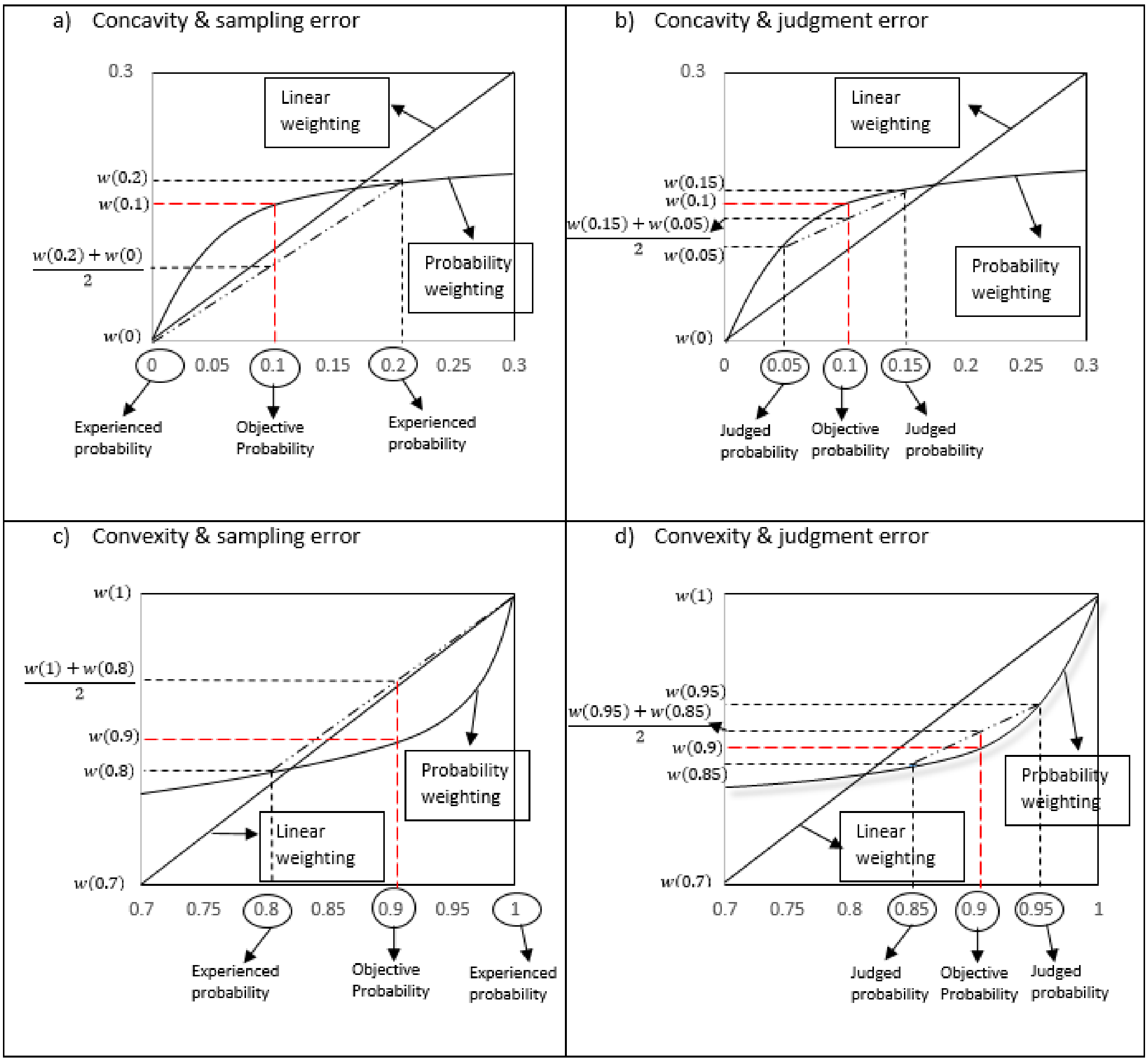
To illustrate, the aforementioned point, assume that all subjects in DFE and DFD have the same probability weighting function depicted in figure 6.2a, which is concave, and overweight 10% probability of a rare and favorable outcome. For the sake of the example, also assume that each subject in DFE draw only 5 times, in which half of the subjects never observe the rare outcome, and the other half observe it once. Therefore, assuming that the subjects do not commit a judgment error, the experienced probabilities will be either 0% or 20%. In this case, aggregating choices over all subjects as commonly done in the DFE literature amounts to averaging the weightings of 10% and 20% rather than weighting the average 10%. This makes the aggregate choice appear as if 10% is underweighted due to concavity whereas in reality it is overweighted (see figure 6.2a).

The same effect, although probably less in size, also applies when there is no sampling error but judgment error. Figure 6.2b illustrates the case where the subjects in DFE accurately observe 10% probability, however, half of them underestimate it as 5% whereas the other half overestimate it as 15%. As a result, the aggregate choice appears as if 10% is weighted less in DFE than in DFD (see figure 6.2b).

By the dual effect, convex probability weighting for large probabilities moves aggregate choices in the direction of overweighting (see figures 6.2c and 6.2d). Together with the concavity around small probabilities, this implies a reversed or attenuated inverse-S at the aggregate level as the DFE-DFD gap also suggests. This theoretical conjecture is indeed indirectly supported by the findings of Rakow et al. (2008). In their yoked design, each subject in the DFE treatment is matched with a subject in the DFD treatment who receives the same sampling information in description format. Thus, equating the heterogeneity of the sampling information across the two treatments, they observe that the DFE-DFD gap is almost completely eliminated.

Our study measures the genuine weightings of probabilities by resolving the possible confounds generated by the aggregation problem. Accordingly, any variations in experienced probabilities are eliminated by matching the sampling experience of each individual subject with the objective probabilities. Subjects are obliged to acquire complete knowledge of probability distributions by sampling all the possible outcomes without replacement, leaving no room for sampling error or ambiguity. Thus, the DFE-DFD comparison turns into a pure comparison of two cases of risk that differ only in terms of information acquisition.

Figure 6.2: Distortions due to aggregation



6.3.4 Underweighting or less overweighting?

Along with the aforementioned issues, the controversy about the DFE-DFD gap is whether it can actually give rise to underweighting, or only less overweighting, of rare outcomes. Rakow and Newell, 2010, pp.6) points out that although the gap implies a relative difference in the weightings of rare outcomes, the evidence on the absolute weightings (over - or under - weighting) is mixed. In particular, the gap often amounts only to a discrepancy in risk attitudes (e.g. different degrees of risk seeking for small probability gains) suggesting a less pronounced overweighting in DFE rather than an absolute underweighting. As a matter of fact, even a reversal in risk attitudes (e.g. risk aversion for small probability gains in DFE as opposed to risk seeking in DFD) may not be sufficient to conclude about the underweighting of rare outcomes under DFE as a concave utility along with an unbiased weighting might also lead to risk aversion.

Our two stage methodology explained in the next section aims to uncover the absolute weighting of probabilities by controlling the utilities. Thus, we clarify the essence of the DFE-DFD gap.

6.4 Method

Our experimental procedure consists of two stages. In the first stage, the utility function of each subject is elicited using the trade-off (TO) method of Wakker and Deneffe (1996). The TO method is a well-established method that is commonly used in studies investigating probability weighting (Abdellaoui 2000; Abdellaoui et al. 2005; Bleichrodt and Pinto 2000; Etchart-Vincent 2004, 2009; Qiu and Steiger 2011). The method basically entails the elicitation of a standard sequence of outcomes that are equally spaced in utility units. The elicitation procedure consists of a series of adaptive indifference relations. For two fixed gauge outcomes G and g, and a selected starting outcome x_o with $x_0 > G > g$, $x_1 > x_0$ is elicited such that the subject is indifferent between prospects $x_{1p}g$ and $x_{0p}G$. Then, x_1 is used as an input to elicit $x_2 > x_1$ such that the subject is indifferent between $x_{2p}g$ and $x_{1p}G$. This procedure is repeated n times in order to obtain the standard sequence $(x_0, , x_n)$ with indifferences $x_{i+1_p}g \sim x_{ip}G$ for $0 \leq i \leq n-1$. Under RDU, these indifferences result in $U(x_1) - U(x_0) = U(x_2) - U(x_1) = ... = U(x_{n-1}) - U(x_n)$ (for the derivation, see Appendix A). A remarkable feature of the TO method is that it elicits these equalities irrespective of what the probability weighting is. Therefore, it is robust against most distortions due to non-expected utility maximization. Once the standard sequence of outcomes has been obtained, we obtain the utility function of each individual by parametrically estimating the power specification $U(x) = x^\alpha$ with $\alpha > 0$ after scaling of $x_i s$ as $x_i = \frac{x_i - x_0}{x_n - x_0}$. We use parametric estimation in order to smooth out errors, and better capture the utility curvature. The parameter α is calculated using an ordinary least squares regression without intercept, $\log(U(x)) = \alpha \log(x) + \epsilon$ where $\epsilon \sim N(0, \sigma^2)$. In the second stage of our procedure, we measure probability weighting using several binary choice questions. The questions are constructed based on the subject-specific outcome sequences obtained from the first stage. Subjects choose between risky prospect $x_{kq}x_j$ and a sure outcome s_q, where x_k and x_j are two distinct elements of the elicited outcome sequence with $x_k > x_j$, and s_q is equal to the certainty equivalent of $x_{kq}x_j$, i.e.

$$s_q = U^{-1}[qU(x_k) + (1-q)U(x_j)] \tag{6.1}$$

That is, s_q would be equivalent to $x_{kq}x_j$ if the subject with the given utility did not weigh probabilities. Hence by construction, the following logical equivalences hold for given preference relations under RDU.

$$x_{kq}x_j \prec s_q \Leftrightarrow w(q) < q \quad (\textit{underweighting}) \tag{6.2}$$

$$x_{kq}x_j \sim s_q \Leftrightarrow w(q) = q \quad (\textit{EU}) \tag{6.3}$$

$$x_{kq}x_j \succ s_q \Leftrightarrow w(q) > q \quad (\textit{overweighting}) \tag{6.4}$$

Because we do not allow indifference in our experiment, each individual choice will reveal either overweighting or underweighting of probability q. Our method makes the deviations from EU observable at the aggregate level. For instance, an overweighting of q can be detected when the majority of subjects choose the risky $x_{kq}x_j$ as in Equation6.4.

Barron and Ursino (2013) also investigates the DFE-DFD gap under risk (their experiment 1) similar to our study by using a different two-stage experimental procedure. Their procedure replicates the well-known DFE-DFD gap. However, it does not make inferences about the over- or under- weighting of rare outcomes under DFE and DFD[4]. Different from Barron and Ursino (2013), our two stage procedure measures the weightings of probabilities under DFE and DFD. Thus, we can draw inferences about actual over - or under - weightings, as well as the DFE-DFD gap.

[4]Their first stage obtains an indifference relation under DFD which implies $w(1-q)U(X) = w(q)U(\$40)$, where the probability q is either 0.1 or 0.2, depending on the treatment, and X was elicited. Their second stage looks at deviations from this indifference under DFE and DFD. Their findings indicate deviations only under DFE, suggesting less weighting of q and/or more weighting of $(1-q)$ under DFE, i.e. $w(1-q)U(X) > w(q)U(\$40)$, consistent with the DFE-DFD gap.

6.5 The Experiment

6.5.1 Subjects and incentives

The experiment was performed at the ESE-EconLab at Erasmus University in 5 group sessions. Subjects were 89 Erasmus University students from various academic disciplines (average age 23 years, 40 female). All subjects were recruited from the pool of subjects who had never participated in any economic experiment in our lab before, to avoid experienced subjects in TO method. We paid each subject a €5 participation fee. In addition, at the end of each session, we randomly selected two subjects who could play out one of their randomly drawn choices for real. The ten subjects who played for real received €60.70 on average. Over the whole experiment, the average payment per subject was €12.37.

6.5.2 Procedure

The experiment was run on computers. Subjects were separated by wooden panels to minimize interaction. To prevent the impact of variations in memory limitations, all subjects were provided with paper and pen in case they wished to take notes. Before they started with the main parts of the experiment, they read the general instructions with detailed information about the payment procedure, the user interface, and the type of questions they would face. The subjects could ask questions at any time during the experiment. The experiment consisted of two successive stages without a break in between. Each stage started with its corresponding instructions, and several training questions to familiarize subjects with the stimuli. Each session took 45 minutes on average, including the payment phase after the experiment.

6.5.3 Stimuli

6.5.3.1 Stage 1: measuring utility

In the first stage of the experiment, a standard sequence of outcomes was elicited using the TO method. We measured x_1, x_2, x_3, x_4,and x_5 from the following five indifferences, with $p = 0.33, G = 17, g = 9$, and $x_0 = 24$:

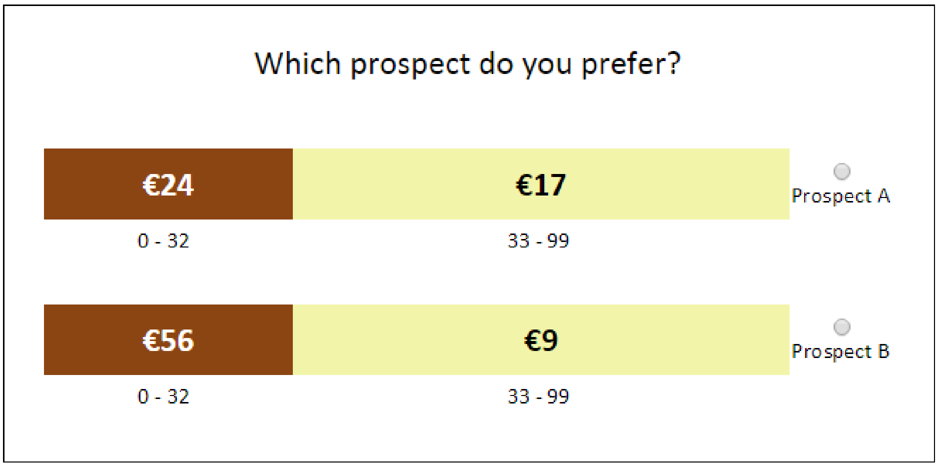
$$24_pG \sim x_{1p}g, x_{1p}G \sim x_{2p}g, x_{2p}G \sim x_{3p}g, x_{3p}G \sim x_{4p}g, x_{4p}G \sim x_{5p}g.$$

Our choice of the fixed parameters p, G, g, x_0 was fine-tuned based on a pilot session so that the elicitation yields a well-spaced outcome sequence for reliable certainty equivalent values of s_q in Equation6.1.

Indifferences were obtained by a bisection method requiring 7 iterations for each x_i. In addition, the last iteration of one randomly chosen x_i was repeated at the end of the stage 1, in order to test the reliability of the indifferences. Hence, subjects answered a total of 36 questions in this part. The bisection iteration procedure is described in Appendix B. The prospects were presented on screen as in Figure 6.3.

In this part, risk was generated by two ten-faced dice each generating one digit of a random number from 00 to 99. The outcome of prospects depended on the result of two dice physically rolled by subjects in case the question was played for real at the end of the experiment.

Figure 6.3: Choice situation in the TO part



6.5.3.2 Stage 2: DFD and DFE

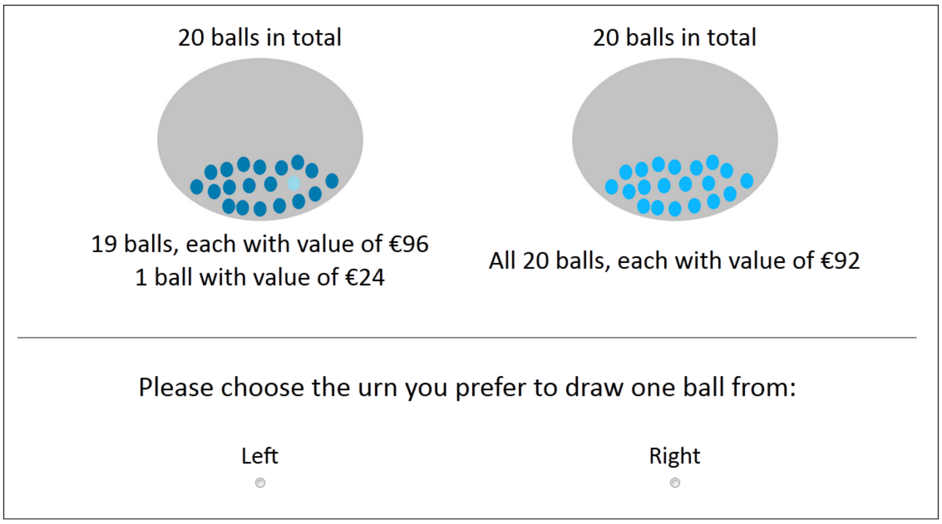
Before the start of the second part, each subject was randomly assigned to one of the two treatments: DFE or DFD. Subjects in both treatments answered 7 subject-specific binary choice questions. Each question entailed a choice between a risky prospect $x_{5q}x_1$

and the safe prospect s_q as described in section 6.4. Note that both x_1 and x_5 were endogenously determined, and varied between subjects (the reason for using x_1, rather than x_0 as the minimum outcome in the risky prospects is explained in the section of Discussion). Values of s_q were always rounded to the nearest integer.

The seven probabilities used for the investigation of probability weighting were 0.05,0.10,0,20,0.50,0.80,0.90 and 0.95. Within each treatment, the orders of the seven questions were counterbalanced. The position of the risky prospect and the safe prospect were also randomized in each question.

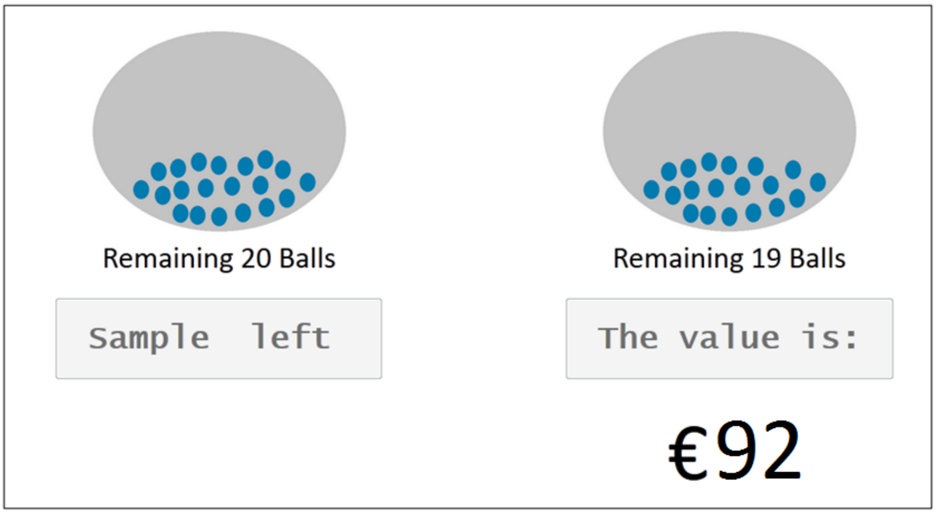
Prospects were represented by Ellsberg-type urns containing 20 balls with different monetary values attached to them. This means that all the aforementioned probabilities were fractions of 20; i.e. 5% is 1 out of 20, 10% is 2 out of 20, etc. The two treatments differed from each other in terms of how the contents of the urns were learnt. Figure 6.4 shows a screen shot of a choice situation for DFD.

Figure 6.4: Choice situation in DFD



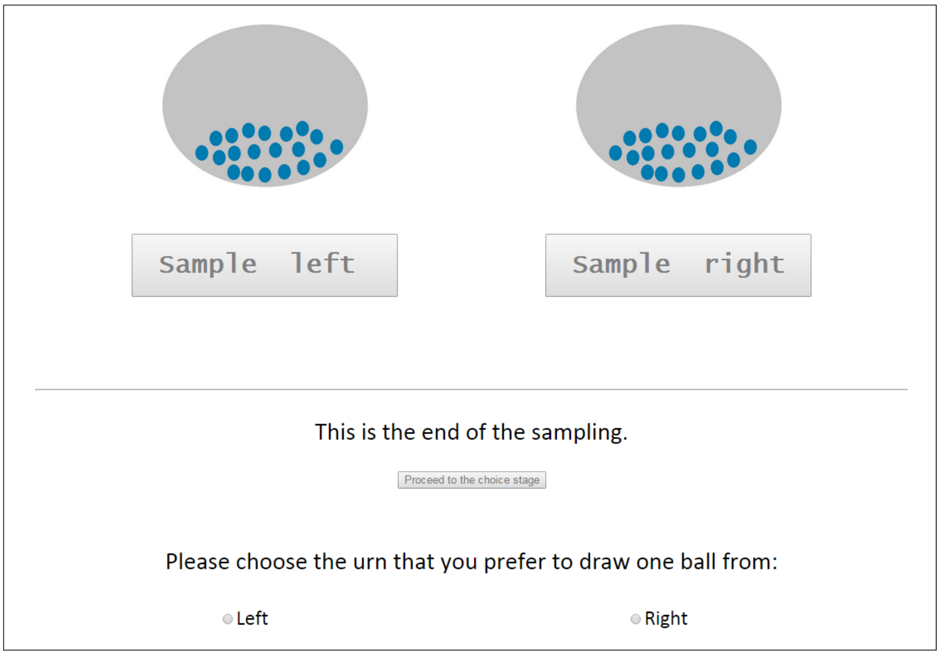
Subjects in the DFE treatment were initially given no information about the contents of the urns except the total number of balls. They could only learn about the outcome compositions of the urns by sampling each and every ball one-by-one without replacement, and observing the monetary values attached. Figure 6.5 shows a screen shot of the sampling phase in the DFE treatment. Subjects sampled balls from urns by clicking "Sample left" or "Sample right" on the screen. Each time, the monetary outcome attached to the ball sampled was shown to the subject for 1.5 seconds, and then disappeared. Subject could sample in their own speed, in whichever order they preferred, and switch as many times as they wanted, but they could only proceed to the choice stage after sampling all the balls in both urns.

Figure 6.5: Choice situation in DFE



The Figure 6.6 shows the screen shot of the choice stage in DFE. In case a question in this part was drawn for the payment at the end of the experiment, the experimenters physically created the relevant urn seen on the screen by filling an opaque urn with 20 ping-pong balls painted to dark blue or light blue, each associated with the payoffs in question (see Figure 6.4). Then, the subject drew a ball from the urn, which determined her payoffs.

Figure 6.6: Choice stage in DFE



Subjects in the DFD treatment faced 21 extra questions following the main set of 7 questions to equalize the length of the two treatments. These extra questions were for another research project.

6.6 Results

6.6.1 Reliability and consistency of utility elicitation

In the TO part, each subject repeated one choice faced in one of the five elicitations. The repeated choice was randomly selected among the last steps of the iterations. Because the subjects were very close to indifference at the last step, this was the strongest test of consistency. Subjects made the same choice in 70.8% of the cases. Reversal rates up to one third are common in the literature (Stott, 2006; Wakker et al., 1994). Especially, if the closeness to indifference is taken into account, our reversal rates are satisfactory. Among the reversed cases, repeated indifferences were higher than the original indifference values in 42.3% of the times, which did not indicate any systematic pattern (p=0.5572, two-sided binomial). Overall, repeated indifference values did not differ from original elicitations (p=0.44, Wilcoxon Sign-rank).

In our data, one subject reached the possible lower bound of x_i's in all 5 cases. Consequently, her standard sequence was not well spaced enough for the estimations of s_q with Equation 6.1 ($x_5 - x_1 = 8$). [5] We excluded this subject from the following analysis. The analysis with this subject included does not alter our conclusions. The same problem was not observed with any other subject.

6.6.2 Utility functions

Table 6.1 gives the descriptive statistics for the elicited outcome sequence. The parameter α of the power utility $u(x) = x^\alpha$ was estimated at the individual level by ordinary least squares regression. The average R^2 over all individual utility estimations was 0.985 which indicated that our estimations fit the data well.

Table 6.1: Descriptive statistics of the elicited outcome sequence (N=88)

	Mean	S.Dev	Min	Median	Max
x_0	24.00	0.00	24.00	24.00	24.00
x_1	60.36	23.48	30.00	58.00	118.00
x_2	90.36	42.58	36.00	80.00	212.00
x_3	125.23	65.89	46.00	102.00	306.00
x_4	164.18	91.13	52.00	134.00	400.00
x_5	204.14	116.25	58.00	160.00	494.00
α	1.05	0.36	0.41	0.99	2.65

The summary statistics for the mean and median α are reported in the last row of Table 6.1. The aggregate data did not deviate from linearity (p=0.92, Wilcoxon sign-rank). Although the mean α suggested slight convexity, this was due to the outliers in our data. Three subjects exhibited extreme convexity with $\alpha > 2$, and the Skewness/Kurtosis test rejected the normality of the distribution of α's (p=0.00). Utilities did not differ across the two treatments (p=0.84, Wilcoxon rank-sum).

Our data suggested slightly more evidence for concavity at the individual level.

[5] Specifically, the resulted estimations, $s_{0.05} = x_1$ and $s_{0.95} = x_5$, made the preference for $x_5 0.05 x_1$ over $s_{0.05}$ and the preference for $s_{0.95}$ over $x_5 0.95 x_1$ trivial because of the domination of the safe or the risky prospect.

Based on the α parameters that were significantly different than 1, 30 subjects (15 in DFE, and 15 in DFD) exhibited concavity ($\alpha < 1$), and 23 subjects (12 in DFE, and 11 in DFD) exhibited convexity ($\alpha > 1$). The proportions of concave and convex utilities did not differ from each other (p=0.41, two-sided binomial).

6.6.3 Probability weighting: DFE vs. DFD

6.6.3.1 Aggregate data

Figure 6.7 shows the proportion of choices in the directions of overweighting and underweighting according to Equation 6.2 and Equation 6.4 in the Method section.

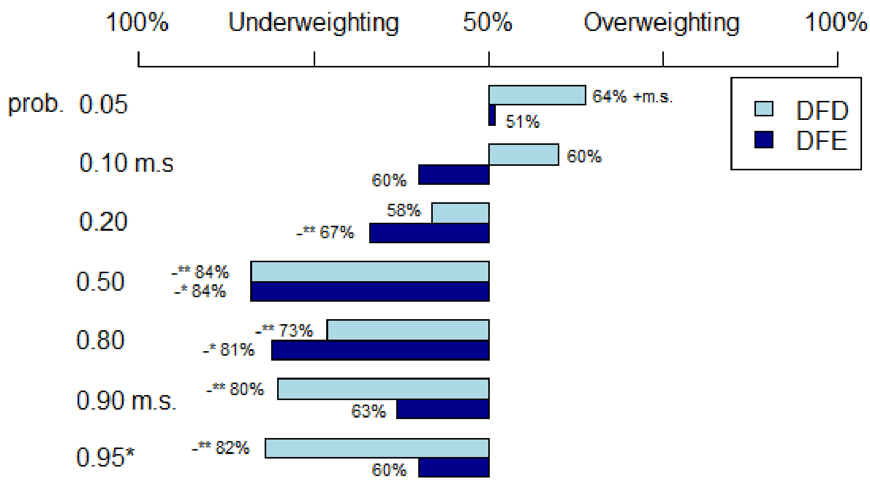
The aggregate choices in the DFD treatment conformed to the inverse-S pattern with strong pessimism: the moderate and the high probabilities ($q \geq 0.50$) were strongly underweighted (p=0.00 for all, two-sided binomial), and the small probability 0.05 was overweighted although this was only marginally significant (p=0.07, two-sided binomial). The deviations from EU in 0.10 and in 0.20 were not significant (p=0.23 and p=0.37 respectively, two-sided binomial). The deviation for these probabilities was in the direction of overweighting for 0.10 and in the direction of underweighting for 0.20.

Turning to the DFE-DFD comparison, the apparent DFE-DFD gap was mainly observed at the extreme probabilities. It was significant at 0.95 (p=0.02, χ^2); and marginally significant at 0.10, and at 0.90 (p=0.06 and p=0.07 respectively, χ^2). The gap was always in the expected direction reducing both the overweighting of the small probabilities and the underweighting of the large probabilities. The gap at the other extreme probability 0.05 was not significant (p=0.20, χ^2), although the trend sugested reduced overweighting in DFE. There was also no apparent DFE-DFD gap at the probabilities in the middle range, $0.20 \leq q \leq 0.80$ (p=0.35, p=0.92 and p=0.37 for q=0.20,0.50, and 0.80 respectively, χ^2).

Our aggregate results did not provide evidence for the reversal of inverse-S in the DFE treatment. Overall, strong underweighting prevailed at the probabilities in the middle range $0.20 \leq q \leq 0.80$ (p=0.00, p=0.03 and p=0.03 for q=0.20,0.50, and 0.8 respectively, two-sided binomial) but there were no significant deviations from unbiased weighting at the extreme probabilities $p \leq 0.10$ and $p \geq 0.90$ (p=1, p=0.22, p=0.13, and p=0.22 for q=0.05,0.10,0.90, and 0.95 respectively, two-sided binomial). Notably, the trend of the gap suggested a reversal in the direction of underweighting in DFE only at 0.10 but this was insignificant.

The findings reported here can be interpreted in terms of likelihood insensitivity and pessimism as discussed before. On the one hand, simultaneous attenuation of both the overweighting of small probabilities and the underweighting of large probabilities suggests increased likelihood sensitivity in the DFE treatment. On the other hand, absence of the significant DFE-DFD gap in the middle range of the probability domain indicates persistent pessimism in both treatments.

Figure 6.7: The DFD-DFE gap



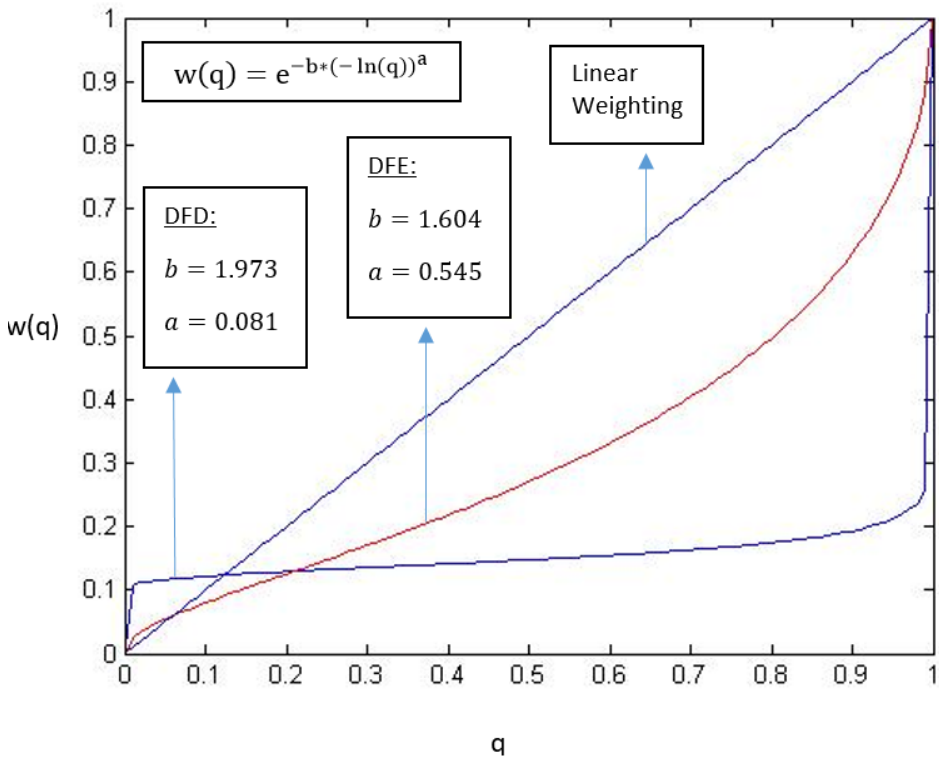
Notes: ms: the gap is marginally significant (at level 10%); *: the gap is significant at level 5%; **: the gap is significant at level 1%; -/+ms: the underweighting/overweighting is marginally significant (at level 10%); -/+*: the underweighting/overweighting is significant at level 5%; -/+**: the underweighting/overweighting is significant at level 1%

An estimation of Prelec's (1998) two-parameter weighting function supports these observations. In this specification given by $w(p) = \exp\left(-\beta(-\ln q)^{\alpha}\right)$, where $0 < \alpha < 1$ and $\beta > 0$, α captures the sensitivity towards changes in probabilities, and β captures the degree of pessimism. We estimated the parameters α and β by the method of maximum likelihood. The probability of choosing the risky prospect was calculated using the stochastic choice rule of Luce (1959) which was also used by Holt and Laury (2002): Pr(choosing risky option)$=\frac{RDU_{risky}^{1/\mu}}{RDU_{risky}^{1/\mu}+RDU_{safe}^{1/\mu}}$, where μ is the noise parameter.

In our case, the rule reduces to Pr(choosing risky option)=$\frac{w(q)^{1/\mu}}{w(q)^{1/\mu}+q^{1/\mu}}$ [6], which implies random choice when $w(q) = q$, consistent with Equation 6.3 in section 6.4 in the Method section. The standard errors were corrected for clustering at subject level.

The estimation results are in Figure 6.8. The probability weighting function for DFD confirms the inverse-S shape as it is steep at both ends, and flat in the middle. The major part of the curve is under the diagonal, signifying strong pessimism. The extreme insensitivity implied by this curve (α=0.081) is discussed in the beginning of the Discussion section.

Figure 6.8: Estimation of Prelec's probability weighting function



The impact of the non-linear probability weighting observed under the DFD treatment is reduced but not reversed under the DFE treatment. Indicating the enhanced likelihood sensitivity in DFE, the sensitivity parameter α differs between the two treatments, although the effect is marginal (p=0.06). Notably, the probability weighting function for DFE also exhibits the inverse S-shape although it is less pronounced than the DFD curve. The small probabilities below 0.05 are still overweighted, and the

[6] Normalizing U(x_1)=0, and U(x_5)=1; $RDU_{risky} = w(q)U(x_5) + (1 - w(q))U(x_1) = w(q)$, and $RDU_{safe} = U(x_q) = q$ by construction

large probabilities are still underweighted but to a lesser extent. The two weighting functions do not differ in terms of pessimism (p=0.34) as both curves lie mostly under the diagonal with almost the same elevation. The same results are replicated using the estimations of Goldstein and Einhorn's (1987) probability weighting functions. The results with this specification are in Appendix C.

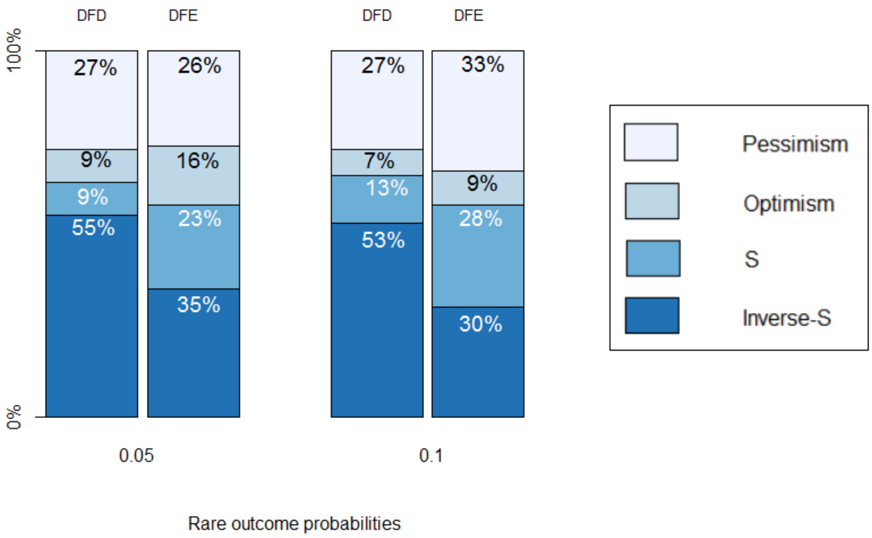
6.6.4 Individual data

We also investigated the shape of probability weighting functions at the individual level. The weighting of probabilities near 0 and 1 are best suited for the investigation of patterns in probability weighting because they reveal the most crucial deviations from linear weighting. We used two separate pairs of probabilities, 0.05 - 0.90, and 0.10 - 0.90, for the individual level analysis. For the pair of 0.05 - 0.95, a probability weighting function was classified as inverse S-shaped if it exhibits $w(0.05) > 0.05$ and $w(0.95) < 0.95$, as S-shaped if it exhibits $w(0.05) < 0.05$ and $w(0.95) > 0.95$, as optimistic if it exhibits $w(0.05) > 0.05$ and $w(0.95) > 0.95$, and as pessimistic if it exhibits $w(0.05) < 0.05$ and $w(0.95) < 0.95$. These inequalities are inferred from Equation 6.2 and Equation 6.4 in the Method section. The classification with the pair of 0.10-0.90 was similar.

Our results are in Figure 6.9. For both probability pairs, inverse S was found to be the dominant pattern in the DFD treatment. The inverse S pattern was significantly more frequent than the second most common pessimistic pattern based on the 0.05 - 0.95 pair (p=0.05, two sided binomial). The difference was marginally significant based on the 0.10 - 0.90 pair (p=0.07, two sided binomial).

Compared to the DFD treatment, the inverse S pattern was found less frequently (p=0.05 for the pair 0.05 - 0.95, and p=0.02 for the pair 0.10 - 0.90, χ^2), and the S pattern was found more frequently in the DFE treatment, although the latter effect was marginal (p=0.07 for the pair 0.05 - 0.95, and p=0.09 for the pair 0.10 - 0.90, χ^2). Despite the differences between the two treatments, there was no clearly dominating pattern in the DFE treatment. Inverse S was still the most common pattern based on the classification with the 0.05 - 0.95 pair, however its proportion did not differ from the proportion of the second most common pessimistic pattern (p=0.56, two sided binomial) or from the proportion of the S-shaped pattern (p=0.33, two sided binomial). Based on the classification with the 0.10 - 0.90 pair, the most common pattern in the DFE treatment was pessimism, however its proportion was not significantly different from the proportion of the inverse-S pattern or from the proportion of the S-pattern (p=1 and p=0.85 respectively, two sided binomial). Hence, the results of the individual analysis confirm those of the aggregate analysis.

Figure 6.9: Classification of probability weighting functions



6.7 Discussion

6.7.1 Experiment design and results

Our adjustment of the sampling paradigm with complete sampling of outcomes aimed to solve confounds generated by sampling error, ambiguity, and the aggregation problem. Camilleri and Newell (2011) argues that forcing subjects to draw large samples can result in other biases due to memory limitations and inattention. One example is the recency effect according to which the more recently observed outcomes are more readily available in memory, and therefore they receive more weight in decisions (Hertwig et al. 2004). To avoid problems related to inattention and memory limitations, we provided our subjects with paper and pen in case they wanted to keep track of the outcomes during the sampling stage in DFE. We observed that more than half of our subjects in DFE treatment took notes. Moreover, our data did not suggest any recency effect. Hence, our results were less like to be driven by other cognitive biases such as limited memory and the recency effect[7].

Our two-stage experimental design controlled for utilities to isolate the impact of probability weighting on risky choices. The utilities were measured under DFD paradigm, and the probability weightings were measured under DFD and DFE. Thus, it is implicit in our design that sampling experience has an impact on probability domain but not on utilities. This conjecture was supported under the sampling paradigm, where the subjects make single decisions without accumulating income (but also see Ludvig and Spetch (2011) with the partial feedback paradigm). In particular, Abdellaoui et al. (2011) measured the utilities under the sampling paradigm of DFE and DFD separately, and did not detect any difference across them.

The first stage of our experiment measured the utilities by using the TO method. One concern about the TO method for utility elicitation is its being adaptive. This means that later stimuli are determined by previous choices. Although the interdependence between different choices may be a problem for the incentive compatibility in theory, all previous studies that investigated the problem found that it does not occur for the TO method (Abdellaoui, 2000; Bleichrodt et al., 2010; Qiu and Steiger, 2010; Schunk and Betsch, 2006; van de Kuilen and Wakker, 2006). Hence, in the terminology of Bardsley et al. (2010), this is only a concern for theoretical incentive compatibility but not for behavioral incentive compatibility (pp. 265). Note that the iterative procedure used for the elicitation of each outcome x_i in the TO method was also adaptive. Our bisection procedure made it difficult for our subjects to understand the adaptive nature of our method by including filler questions in the iteration process. Our data did not show any evidence of strategic choices. See appendix B for details.

The second stage measured probability weightings by using binary choice questions constructed based on the estimated power utilities in the first stage. In these questions,

[7]Observing the rare outcome(s) in the first or in the second half of the sequence did not have an impact on risky choices (p=0.84, p=0.87, p=0.85, and p=0.15 for q=0.05,0.10,0.90, and 0.95 respectively, chi^2).

we used the elicited x_1 as the minimum outcome of the risky prospects to avoid problems related to the extreme behavior of power utility near its origin (Wakker 2008), i.e. x_0 in our design. In particular, for $\alpha<1$, the slope of the power utility converges to infinity as x tends to the origin. This implies extreme risk aversion near the origin. Similarly, $\alpha>1$ implies extreme risk seeking near the origin. The replication of the common inverse S pattern in our DFD treatment confirmed the robustness of our design.

Our results indicated likelihood insensitivity in both DFD and DFE but to a significantly lesser degree in the latter. The extreme likelihood insensitivity implied by the estimations of Prelec's weighting function in the DFD treatment might be surprising, regarding more moderate estimations reported in the previous studies under DFD. However, one important difference of our study is the use of binary choices - rather than certainty equivalents - as the method of elicitation measurement. As the accounts of the preference reversal phenomenon suggest, the prominence of the probability domain in binary choices might possibly enhance the impact of nonlinear probability weighting on risk (see Tversky et al. (1990) on prominence hypothesisthe prominence hypothesis). Our individual level analysis confirmed the extreme insensitivity in the DFD treatment as 55% and 53% of the probability weighting functions were classified as inverse S-shaped based on 0.05 - 0.95 and 0.10 - 0.90 pairs respectively, whereas only 9% and 13% were classified as S-shaped (likelihood sensitivity).

Our finding of the attenuated inverse S in DFE was also supported by a very recent study (Kopsacheilis, 2016, April). In the same vein as our study, Kopsacheilis controlled for the sampling error, ambiguity and memory effects by introducing several treatments manipulating these effects. His findings mainly indicated that the sampling error was the most important component of the gap, and the error-free sampling experience led only to a reduced overweighting of small probabilities.

Our findings did not suggest any gap for moderate probabilities. Although outcomes with moderate probabilities, such as 50%, have received little attention in DFE literature, Lejarraga et al. (2016) and Ludvig and Spetch (2011) suggested that the DFE-DFD gap also extends beyond rare outcomes. There are two possible explanations for this discrepancy between the findings. First, while our study uses an adjustment of the sampling paradigm, Lejarraga et al. (2016) uses the original sampling paradigm in which the experienced probabilities is uncertain (ambiguity), and Ludvig and Spetch (2011) uses the feedback paradigm in which subjects accumulate income. Second, while our study focuses on the gain domain, Ludvig and Spetch (2011) considers both the gain and the loss domain, and Lejarraga et al. (2016) considers the loss domain (in monetary and health outcomes). The true impact of experience in the loss domain awaits future research.

6.7.2 The impact of learning experience

The experimental results show that the deviations from rational weighting of probabilities, which are often associated with cognitive limitations, diminish as the probabilistic information is acquired through sequential sampling. Learning from experience enhances sensitivity to changes in relative frequencies but does not impact the attractiveness of risky prospects. The apparent discrepancy between the two informationally-identical treatments signifies the distinct mental processing and representation of the event frequencies in memory resulted from experience. See Estes (1976) for a detailed cognitive account of probability learning.

Our findings support Plott's (1996) Discovered Preference Hypothesis which states that so-called anomalies in human choice diminish or even disappear with proper learning opportunities and familiarity with decision problems. Accordingly, learning from experience may correct the misunderstandings about the meaning of probabilities, and enhance EU maximization. Plott's account is also supported by other previous studies investigating the impact of learning experience. Gottlieb et al. (2007), Hilbig and Glockner (2011), and Humphrey (2006) report reduced probability weighting with different variants of the sampling paradigm. Erev et al. (2015, problems 1,2,7 to 11), Jessup et al. (2008), van de Kuilen and Wakker (2006) and van de Kuilen (2007) report significant convergence to EU maximization under risk in repeated choice settings, when immediate feedback after each choice is available but not when it is unavailable. This suggests the distinct impact of experience in repeated choice settings. Another recent study by Yechiam et al. (2015) reports strong underweighting of very small probabilities, such as 0.005, over 200 to 400 trials with accumulating payoffs. The results of

Yechiam et al. call for further investigation of the weightings of very small probabilities when the income effect is absent, and the utilities are controlled for.

6.8 Conclusion

This paper clarifies the controversy about the DFE-DFD gap. Our strictly controlled sampling paradigm isolates the impact of the sampling experience from other confoundings, and the rigorous two stage design reveals the exact weighting of probabilities under DFE. The experimental findings support the DFE-DFD gap. However, the gap does not amount to a reversal of the inverse S-shaped probability weighting and there is no actual underweighting of rare and extreme outcomes in DFE. Our findings illustrate the importance of the learning experience in reducing irrationalities. Decisions from experience do not reverse an irrationality into another irrationality but rather the cognitive impairment of likelihood insensitivity. Black swans are not ignored under DFE.

6.9 Appendix A. Derivation of the standard sequence of outcomes in TO method

Under RDU, indifferences $x_{i+1_p}g \sim x_{i_p}G$ imply $w(p)U(x_{i+1}) + (1 - w(p))U(g) = w(p)U(x_i)+(1-w(p))U(G)$. A rearrangement of this equation shows $U(x_{i+1})-U(x_i) = \frac{1-w(p)}{w(p)}[U(G) - U(g)]$ for all $0 \leq i \leq n - 1$. Because the right hand side of the equation is fixed by the design, the indifferences result in $U(x_1) - U(x_0) = U(x_2) - U(x_1) = ... = U(x_{n-1}) - U(x_n)$.

6.10 Appendix B. Bisection Procedure

The iteration process serves to measure x_1, x_2, x_3, x_4, and x_5 from the following indifferences, with $p = 0.33, G = 17, g = 9, x_0 = 24$:

$$x_{0p}G \sim x_{1p}g, x_{1p}G \sim x_{2p}g, x_{2p}G \sim x_{3p}g, x_{3p}G \sim x_{4p}g, x_{4p}G \sim x_{5p}g.$$

For each x_i, it took five choice questions to reach the indifference point. Subjects always chose between two prospects: $x_{ip}g$ and $x_{i-1_p}G$ for $i = 1, ..., 5$. The procedure was as follows.

1. The initial value of x_i was determined as $x_{i-1} + 4(G - g) = x_{i-1} + 32$.

2. x_i was increased by a given step size when $x_{i-1_p}G$ was chosen over $x_{ip}g$, and decreased when $x_{ip}g$ was chosen over $x_{i-1_p}G$ as long as $x_i > x_{i-1}$. In case $x_i \leq x_{i-1}$, x_i was increased in order to ensure outcome monotonicity.

3. The initial step was $4(G - g) = 32$. The step sizes were halved after each choice.

4. The indifference point was reached after five choices.

5. The largest possible value of x_i was $x_{i-1} + 32 + 32 + 16 + 8 + 4 + 2 = x_{i-1} + 94$.

6. The smallest possible value of x_i was $x_{i-1} + 32 - 32 + 16 - 8 - 4 - 2 = x_{i-1} + 2$. The fourth term on the left hand side $(+16)$ ensured the outcome monotonicity (see point 2).

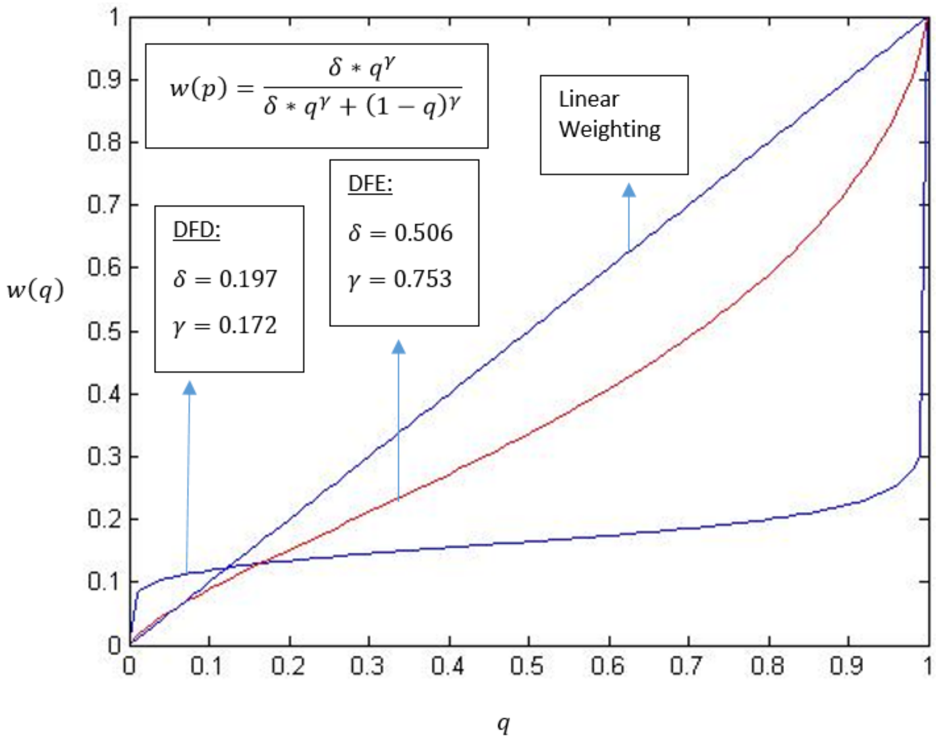
One concern for the TO method and the bisection iteration process is the incentive compatibility due to the adaptive design. A subject who is fully aware of the adaptive design can strategically drive the value x_i upwards by pretending to be extremely risk averse in the bisection questions. In this way, he or she can increase the expected values of prospects in the subsequent questions for the elicitation ofx_{i+1}. To make it more difficult for our subjects to fully grasp the process, we included two filler questions in the iteration process of each x_i. The two filler choices were after the first and the third choice questions for every x_i. In these questions, x_i was changed in the direction that is opposite to the changes described in point 2 above. These questions had no further impact on the flow of the procedure.

Our data did not suggest any strategic behavior. While an awareness of the adaptive design from the outset is fairly unlikely, learning during the experiment would lead to increasing distances between x_is. This means that a systematic learning of the strategic choice during the experiment would give us larger distances between x_5 and x_4 than between x_1 and x_0. On the contrary, the median distances in our data were 26 and 34 respectively, and did not differ significantly (p=0.54, Wilcoxon sign-rank).

6.11 Appendix C: Estimations of Goldstein and Einhorn's (1987) probability weighting function

Goldstein and Einhorn's (1987) probability weighting function is given by $w(q) = \frac{\delta q^{\gamma}}{(\delta q^{\gamma}+(1-q)^{\gamma}}$. In this specification γ captures the likelihood insensitivity whereas δ is the anti-pessimism parameter. The sensitivity parameter γ differed between treatments, although the effect was marginally significant ($p = 0.07$). The anti-pessimism parameter δ did not differ between the treatments ($p = 0.19$). These results are shown in figure 6.10 below.

Figure 6.10: Estimation of G & E's probability weighting function



Chapter 7

General Conclusions

This dissertation investigated the stability, and adaptivity of decision making over time and under risk. Chapter 2 and 3 addressed the stability side of preference. Chapter 2 introduced a new method to measure the temporal discounting of money. Chapter 3 tested reduction invariance, and confirmed the validity of Prelec's compound-invariant probability weighting function. The latter three chapters continues to explore the adaptivity side where preferences were studied in different outcomes and under various context. Chapter 4 compared the deviations from constant discounting for health and money. Chapter 5 elicited people's risk attitudes when using cash vs. numbers. Chapter 6 discussed if people have completely different risk preferences when making decisions from experience, and provided new evidence to the DFD-DFE gap literature.

Chapter 2 presented the direct method to measure discounting that requires no knowledge of utility. Although all existing measurements of money discounting have used discrete outcomes, real-life decisions often involve flow outcomes, such as salary payments, pension saving plans, and mortgage debt repayments. In such contexts, the DM is more natural than discrete methods. However, we have to admit that our method is less useful for single-outcome decisions.

Chapter 3 followed up on Luce (2001) and performed an experimental test of reduction invariance. Our data support reduction invariance, indicating that the behavioral foundation of Prelec's probability weighting function is empirically valid. A byproduct of our study is that we also tested reduction of compound gambles, which is often considered a feature of rational choice. Most subjects (60%) behaved according to it.

The subjects who deviated from it, deviated overwhelmingly in the direction of higher certainty equivalents for the compound gambles than for the corresponding simple gambles.

Chapter 4 assessed the severity of departures from constant impatience, and quantified the degree of decreasing impatience for both health and money. Knowing whether time preferences are the same for those two domains is important for both research and policy. The results indicated that most subjects deviated from constant impatience and were decreasingly impatient for both health and money. Subjects deviated more from constant discounting for health than for money. This domain-dependence of discounting suggests that evidence on time preferences for money has only limited validity for health. It is inappropriate for policy makers to adopt the time preference in money to health directly. In addition, about one third of our subjects exhibited increasing impatience, suggesting that their willingness to wait decreases as time passes by. This finding cannot be incorporated by most discounting models.

Chapter 5 demonstrated the malleability of risk attitudes. By just presenting lottery outcomes in cash instead of numbers, people deviated more from rationality: they were less sensitive towards changes in likelihoods. Our results also provided an explanation for credit card premium: when the value of the product is uncertain, customers using cash are more risk averse than those who use credit cards, therefore holding credit cards makes you spend more.

Chapter 6 investigated the DFE-DFD gap by solving four problems in the previous studies: sampling error due to limited search, ambiguity due to sampling with replacement, distorted aggregate level results caused by aggregation of heterogeneous individual data, and unclear probability weighting because of unadjusted utility functions. Our results did not support the prevalent finding of S-shape probability weighting in DFE research: the gap still existed while it did not amount to a reversal of the inverse S-shaped probability weighting. We signify the importance of the learning experience in the reduction of irrationalities. Decision from experience did not reverse an irrationality into another irrationality but rather reduced the irrationality in the presence of proper learning experience that was free from sampling biases and ambiguity.

Bibliography

Abdellaoui, M. (2000). Parameter-free elicitation of utility and probability weighting functions. *Management Science*, 46(11):1497–1512.

Abdellaoui, M., Attema, A. E., and Bleichrodt, H. (2010). Intertemporal Tradeoffs for Gains and Losses: An Experimental Measurement of Discounted Utility*. *The Economic Journal*, 120(545):845–866.

Abdellaoui, M., Bleichrodt, H., and l'Haridon, O. (2013). Sign-dependence in intertemporal choice. *Journal of Risk and Uncertainty*, 47(3):225–253.

Abdellaoui, M., Klibanoff, P., and Placido, L. (2015). Experiments on Compound Risk in Relation to Simple Risk and to Ambiguity. *Management Science*, 61(6):1306–1322.

Abdellaoui, M., L'Haridon, O., and Paraschiv, C. (2011). Experienced vs. Described Uncertainty: Do We Need Two Prospect Theory Specifications? *Management Science*, 57(10):1879–1895.

Aczel, J. (1966). *Lectures on Functional Equations and Their Applications*. New York: Academic Press.

Aczel, J. and Luce, R. D. (2007). A behavioral condition for Prelec's weighting function on the positive line without assuming. *Journal of Mathematical Psychology*, 51(2):126–129.

Andersen, S., Harrison, G. W., Lau, M. I., and Rutstrom, E. E. (2008). Eliciting Risk and Time Preferences. *Econometrica*, 76(3):583–618.

Andreoni, J. and Sprenger, C. (2012a). Estimating Time Preferences from Convex Budgets. *The American Economic Review*, 102(7).

Andreoni, J. and Sprenger, C. (2012b). Risk preferences are not time preferences. *The American Economic Review*, 102(7).

Attema, A. E. (2012). Developments in time preference and their implications for medical decision making. *Journal of the Operational Research Society*, 63(10):1388–1399.

Attema, A. E., Bleichrodt, H., Rohde, K. I. M., and Wakker, P. P. (2010). Time-Tradeoff Sequences for Analyzing Discounting and Time Inconsistency. *Management Science*, 56(11):2015–2030.

Attema, A. E., Bleichrodt, H., and Wakker, P. P. (2012). A direct method for measuring discounting and qalys more easily and reliably. *Medical Decision Making*, 32(4):583–593.

Attema, A. E., Brouwer, W. B. F., and l'Haridon, O. (2013). Prospect theory in the health domain: A quantitative assessment. *Journal of Health Economics*, 32(6):1057–1065.

Bagnall, J., Bounie, D., Huynh, K. P., Kosse, A., Schmidt, T., Schuh, S. D., and Stix, H. (2014). Consumer cash usage: A cross-country comparison with payment diary survey data.

Baillon, A. (2008). Eliciting Subjective Probabilities Through Exchangeable Events: An Advantage and a Limitation. *Decision Analysis*, 5(2):76–87.

Bar-Hillel, M. (1973). On the subjective probability of compound events. *Organizational Behavior and Human Performance*, 9(3):396–406.

Bardsley, N., Cubitt, R., Loomes, G., Moffatt, P., Starmer, C., and Sugden, R. (2010). *Experimental Economics: Rethinking the Rules*. Princeton University Press.

Barron, G. and Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making*, 16(3):215–233.

Barron, G. and Ursino, G. (2013). Underweighting rare events in experience based decisions: Beyond sample error. *Journal of Economic Psychology*, 39:278–286.

Barron, G. and Yechiam, E. (2009). The coexistence of overestimation and underweighting of rare events and the contingent recency effect. *Judgment and Decision Making*, 4(6):447.

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

Bernasconi, M. and Loomes, G. (1992). Failures of the reduction principle in an Ellsberg-type problem. *Theory and Decision*, 32(1):77–100.

Bleichrodt, H., Cillo, A., and Diecidue, E. (2010). A Quantitative Measurement of Regret Theory. *Management Science*, 56(1):161–175.

Bleichrodt, H. and Johannesson, M. (2001). Time Preference for Health: A Test of Stationarity versus Decreasing Timing Aversion. *Journal of Mathematical Psychology*, 45(2):265–282.

Bleichrodt, H., Kothiyal, A., Prelec, D., and Wakker, P. P. (2013). Compound invariance implies prospect theory for simple prospects. *Journal of Mathematical Psychology*, 57(3âĂŞ4):68–77.

Bleichrodt, H. and Pinto, J. L. (2000). A Parameter-Free Elicitation of the Probability Weighting Function in Medical Decision Analysis. *Management Science*, 46(11):1485–1496.

Bleichrodt, H. and Pinto, J. L. (2002). Loss aversion and scale compatibility in two-attribute trade-offs. *Journal of Mathematical Psychology*, 46(3):315–337.

Bleichrodt, H., Rohde, K. I. M., and Wakker, P. P. (2009). Non-hyperbolic time inconsistency. *Games and Economic Behavior*, 66(1):27–38.

Booij, A. S., Praag, B. M. S. v., and Kuilen, G. v. d. (2010). A parametric analysis of prospect theoryâĂŹs functionals for the general population. *Theory and Decision*, 68(1-2):115–148.

Bostic, R., Herrnstein, R. J., and Luce, R. D. (1990). The effect on the preference-reversal phenomenon of using choice indifferences. *Journal of Economic Behavior & Organization*, 13(2):193–212.

Broome, J. (1991). *Weighing Goods: Equality, Uncertainty and Time*. Wiley-Blackwell.

Bruhin, A., Fehr-Duda, H., and Epper, T. (2010). Risk and Rationality: Uncovering Heterogeneity in Probability Distortion. *Econometrica*, 78(4):1375–1412.

Budescu, D. V. and Fischer, I. (2001). The same but different: an empirical investigation of the reducibility principle. *Journal of Behavioral Decision Making*, 14(3):187–206.

Cairns, J. and Van der Pol, M. (1997). Constant and decreasing timing aversion for saving lives. *Social Science & Medicine*, 45(11):1653–1659.

Cairns, J. A. (1992). Health, wealth and time preference. *Project Appraisal*, 7(1):31–40.

Cairns, J. A. (1994). Valuing future benefits. *Health Economics*, 3(4):221–229.

Camerer, C. (1995). Individual Decision Making. In Kagel, J. H. and Roth, A. E., editors, *Handbook of Experimental Economics*, pages 587–703. Princeton University Press, Princeton, NJ.

Camilleri, A. R. and Newell, B. R. (2009). The role of representation in experience-based choice. *Judgment and Decision Making*, 4(7):518.

Camilleri, A. R. and Newell, B. R. (2011). When and why rare events are underweighted: A direct comparison of the sampling, partial feedback, full feedback and description choice paradigms. *Psychonomic Bulletin & Review*, 18(2):377–384.

Chapman, G. B. (1996a). Expectations and Preferences for Sequences of Health and Money. *Organizational Behavior and Human Decision Processes*, 67(1):59–75.

Chapman, G. B. (1996b). Temporal discounting and utility for health and money. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(3):771–791.

Chapman, G. B. and Elstein, A. S. (1995). Valuing the Future Temporal Discounting of Health and Money. *Medical Decision Making*, 15(4):373–386.

Chapman, G. B., Nelson, R., and Hier, D. B. (1999). Familiarity and time preferences: Decision making about treatments for migraine headaches and Crohn's disease. *Journal of Experimental Psychology: Applied*, 5(1):17–34.

Chechile, R. A. and Barch, D. H. (2013). Using logarithmic derivative functions for assessing the risky weighting function for binary gambles. *Journal of Mathematical Psychology*, 57:15–28.

Chesson, H. W. and Viscusi, W. K. (2003). Commonalities in Time and Ambiguity Aversion for Long-Term Risks*. *Theory and Decision*, 54(1):57–71.

Cho, Y., Duncan, R., and von Winterfeldt, D. (1994). Tests of assumptions about the joint receipt of gambles in rank- and sign-dependent utility theory. *Journal of Experimental Psychology: Human Perception and Performance*, 20(5):931–943.

Cho, Y. and Luce, R. D. (1995). Tests of Hypotheses about Certainty Equivalents and Joint Receipt of Gambles. *Organizational Behavior and Human Decision Processes*, 64(3):229–248.

Chung, N.-K., Winterfeldt, D. v., and Luce, R. D. (1994). An Experimental Test of Event Commutativity in Decision Making Under Uncertainty. *Psychological Science*, 5(6):394–400.

Cropper, M. L., Aydede, S. K., and Portney, P. R. (1994). Preferences for life saving programs: how the public discounts time and age. *Journal of Risk and Uncertainty*, 8(3):243–265.

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

Currim, I. S. and Sarin, R. K. (1989). Prospect Versus Utility. *Management Science*, 35(1):22–41.

Davis-Stober, C. P. (2009). Analysis of multinomial models under inequality constraints: Applications to measurement theory. *Journal of Mathematical Psychology*, 53(1):1–13.

Davis-Stober, C. P., Brown, N., and Cavagnaro, D. R. (2015). Individual differences in the algebraic structure of preferences. *Journal of Mathematical Psychology*, 66:70–82.

DellaVigna, S. (2009). Psychology and Economics: Evidence from the Field. *Journal of Economic Literature*, 47(2):315–72.

Denrell, J. and March, J. (2001). Adaptation as Information Restriction: The Hot Stove Effect. *Organization Science*, 12(5):523–538.

Diamond, P. and Köszegi, B. (2003). Quasi-hyperbolic discounting and retirement. *Journal of Public Economics*, 87(9-10):1839–1872.

Diecidue, E., Schmidt, U., and Zank, H. (2009). Parametric weighting functions. *Journal of Economic Theory*, 144(3):1102–1118.

Dolan, P. and Kahneman, D. (2008). Interpretations Of Utility And Their Implications For The Valuation Of Health. *The Economic Journal*, 118(525):215–234.

Ebert, J. E. J. and Prelec, D. (2007). The Fragility of Time: Time-Insensitivity and Valuation of the Near and Far Future. *Management Science*, 53(9):1423–1438.

Epper, T. and Fehr-Duda, H. (2015). Balancing on a Budget Line: Comment on Andreoni and Sprenger's sRisk Preferences Are Not Time Preferences. *American Economic Review*, 105(7):2261–2271.

Epper, T., Fehr-Duda, H., and Bruhin, A. (2011). Viewing the future through a warped lens: Why uncertainty generates hyperbolic discounting. *Journal of Risk and Uncertainty*, 43(3):169–203.

Erev, I., Glozman, I., and Hertwig, R. (2008). What impacts the impact of rare events. *Journal of Risk and Uncertainty*, 36(2):153–177.

Erev, I., Plonsky, O., and Ert, E. (2015). From Anomalies to Forecasts: A Choice Prediction Competition for Decisions under Risk and Ambiguity. *mimeo*.

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological bulletin*, 53(2):134.

Estes, W. K. (1976). The cognitive side of probability learning. *Psychological Review*, 83(1):37.

Estes, W. K. (2002). Traps in the route to models of memory and decision. *Psychonomic Bulletin & Review*, 9(1):3–25.

Etchart-Vincent, N. (2004). Is Probability Weighting Sensitive to the Magnitude of Consequences? An Experimental Investigation on Losses. *Journal of Risk and Uncertainty*, 28(3):217–235.

Fang, H. and Wang, Y. (2015). Estimating Dynamic Discrete Choice Models with Hyperbolic Discounting, with an Application to Mammography Decisions. *International Economic Review*, 56(2):565–596.

Fehr-Duda, H., Gennaro, M. d., and Schubert, R. (2006). Gender, Financial Risk, and Probability Weights. *Theory and Decision*, 60(2-3):283–313.

Feinberg, R. A. (1986). Credit cards as spending facilitating stimuli: A conditioning interpretation. *Journal of consumer research*, pages 348–356.

Fox, C. R. and Hadar, L. (2006). " Decisions from experience"= sampling error+ prospect theory: Reconsidering Hertwig, Barron, Weber & Erev (2004). *Judgment and Decision Making*, 1(2):159.

Fox, C. R. and Poldrack, R. A. (2014). Prospect theory and the brain. *Handbook of neuroeconomics(2nd Ed.)*, pages 533–567.

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

Friedman, Z. G. (2005). Evaluating risk in sequential events: Violations in expected utility theory. *Washington University Undergraduate Research Digest*, 1:29–39.

Gerber, A. and Rohde, K. I. M. (2010). Risk and preference reversals in intertemporal choice. *Journal of Economic Behavior & Organization*, 76(3):654–668.

Gerber, A. and Rohde, K. I. M. (2015). Eliciting discount functions when baseline consumption changes over time. *Journal of Economic Behavior & Organization*, 116:56–64.

Goldstein, W. M. and Einhorn, H. J. (1987). Expression theory and the preference reversal phenomena. *Psychological Review*, 94(2):236–254.

Gonzalez, R. and Wu, G. (1999). On the Shape of the Probability Weighting Function. *Cognitive Psychology*, 38(1):129–166.

Gottlieb, D. A., Weiss, T., and Chapman, G. B. (2007). The Format in Which Uncertainty Information Is Presented Affects Decision Biases. *Psychological Science*, 18(3):240–246.

Gruber, J. and Köszegi, B. (2001). Is Addiction "Rational"? Theory and Evidence. *The Quarterly Journal of Economics*, 116(4):1261–1303.

Gruber, J. and Köszegi, B. (2004). Tax incidence when individuals are time-inconsistent: the case of cigarette excise taxes. *Journal of Public Economics*, 88(9-10):1959–1987.

Gul, F. (1991). A theory of disappointment aversion. *Econometrica: Journal of the Econometric Society*, pages 667–686.

Hadar, L. and Fox, C. R. (2009). Information asymmetry in decision from description versus decision from experience. *Judgment and Decision Making*, 4(4):317.

Halevy, Y. (2015). Time Consistency: Stationarity and Time Invariance. *Econometrica*, 83(1):335–352.

Hardisty, D. J. and Weber, E. U. (2009). Discounting future green: Money versus the environment. *Journal of Experimental Psychology: General*, 138(3):329–340.

Harvey, C. (1995). Proportional Discounting of Future Costs and Benefits. *Mathematics of Operations Research*, 20(2):381–399.

Harvey, C. M. (1986). Value functions for infinite-period planning. *Management Science*, 32(9):1123–1139.

Hastie, T., Tibshirani, R., Friedman, J., and Franklin, J. (2008). *The elements of statistical learning: data mining, inference and prediction (2nd ed.)*, volume 27. Springer, Berlin: Springer series in statistics.

Hau, R., Pleskac, T. J., and Hertwig, R. (2010). Decisions from experience and statistical probabilities: Why they trigger different choices than a priori probabilities. *Journal of Behavioral Decision Making*, 23(1):48–68.

Hau, R., Pleskac, T. J., Kiefer, J., and Hertwig, R. (2008). The descriptionâĂŞexperience gap in risky choice: the role of sample size and experienced probabilities. *Journal of Behavioral Decision Making*, 21(5):493–518.

Hertwig, R., Barron, G., Weber, E. U., and Erev, I. (2004). Decisions from Experience and the Effect of Rare Events in Risky Choice. *Psychological Science*, 15(8):534–539.

Hertwig, R. and Erev, I. (2009). The description-experience gap in risky choice. *Trends in Cognitive Sciences*, 13(12):517–523.

Hilbig, B. E. and Glockner, A. (2011). Yes, they can! Appropriate weighting of small probabilities as a function of information acquisition. *Acta Psychologica*, 138(3):390–396.

Hirschman, E. C. (1979). Differences in consumer purchase behavior by credit card payment system. *Journal of Consumer Research*, pages 58–66.

Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *American economic review*, 92(5):1644–1655.

Humphrey, S. J. (2006). Does Learning Diminish Violations of Independence, Coalescing and Monotonicity? *Theory and Decision*, 61(2):93–128.

Ikeda, S., Kang, M.-I., and Ohtake, F. (2010). Hyperbolic discounting, the sign effect, and the body mass index. *Journal of Health Economics*, 29(2):268–284.

Jeffreys, S. H. (1961). *The Theory of Probability*. New York: Oxford University Press.

Jessup, R. K., Bishara, A. J., and Busemeyer, J. R. (2008). Feedback Produces Divergence From Prospect Theory in Descriptive Choice. *Psychological Science*, 19(10):1015–1022.

Kahn, B. E. and Sarin, R. K. (1988). Modeling ambiguity in decisions under uncertainty. *Journal of Consumer Research*, pages 265–272.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Karmarkar, U. S. (1978). Subjectively weighted utility: A descriptive extension of the expected utility model. *Organizational Behavior and Human Performance*, 21(1):61–72.

Keeney, R. L. and Raiffa, H. (1976). *Decisions with multiple objectives: preferences and value trade-offs*. Cambridge university press.

Keller, L. R. (1985). Testing of the "reduction of compound alternatives" principle. *Omega*, 13(4):349–358.

Kemel, E. and Travers, M. (2015). Comparing attitudes toward time and toward money in experience-based decisions. *Theory and Decision*, 80(1):71–100.

Keren, G. and Wagenaar, W. A. (1987). Violation of utility theory in unique and repeated gambles. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(3):387.

Khwaja, A., Silverman, D., and Sloan, F. (2007). Time preference, time discounting, and smoking decisions. *Journal of Health Economics*, 26(5):927–949.

Koopmans, T. C. (1960). Stationary Ordinal Utility and Impatience. *Econometrica*, 28(2):287–309.

Kopsacheilis, O. (2016). Information about risky decisions: Decomposing and reinterpreting the description-experience gap. *Poster session presented at the NIBS conference, the University of East Anglia; 5, April, 2016; Norwich, UK.*

Kouchaki, M., Smith-Crowe, K., Brief, A. P., and Sousa, C. (2013). Seeing green: Mere exposure to money triggers a business decision frame and unethical outcomes. *Organizational Behavior and Human Decision Processes*, 121(1):53–61.

Krawczyk, M. W. (2015). Probability weighting in different domains: The role of affect, fungibility, and stakes. *Journal of Economic Psychology.*

Laibson, D. (1997). Golden Eggs and Hyperbolic Discounting. *The Quarterly Journal of Economics*, 112(2):443–478.

Lattimore, P. K., Baker, J. R., and Witte, A. D. (1992). The influence of probability on risky choice: A parametric examination. *Journal of Economic Behavior & Organization*, 17(3):377–400.

Lejarraga, T., Pachur, T., Frey, R., and Hertwig, R. (2016). Decisions from Experience: From Monetary to Medical Gambles. *Journal of Behavioral Decision Making*, 29(1):67–77.

Loewenstein, G. (1987). Anticipation and the valuation of delayed consumption. *The Economic Journal*, pages 666–684.

Loewenstein, G. and Prelec, D. (1991). Negative time preference. *The American Economic Review*, pages 347–352.

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

Loewenstein, G. F. and Prelec, D. (1993). Preferences for sequences of outcomes. *Psychological Review*, 100(1):91–108.

Loewenstein, G. F., Weber, E. U., Hsee, C. K., and Welch, N. (2001). Risk as feelings. *Psychological bulletin*, 127(2):267.

Lopes, L. L. (1982). Doing the impossible: A note on induction and the experience of randomness. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8(6):626.

Luce, D. (2002). A psychophysical theory of intensity proportions, joint presentations, and matches. *Psychological Review*, 109(3):520–532.

Luce, D. (2004). Symmetric and Asymmetric Matching of Joint Presentations. *Psychological Review*, 111(2):446–454.

Luce, R. D. (1959). On the possible psychophysical laws. *Psychological review*, 66(2):81.

Luce, R. D. (1991). Rank- and sign-dependent linear utility models for binary gambles. *Journal of Economic Theory*, 53(1):75–100.

Luce, R. D. (1996). When Four Distinct Ways to Measure Utility Are the Same. *Journal of Mathematical Psychology*, 40(4):297–317.

Luce, R. D. (2000). *Utility of gains and losses: Measurement-theoretical and experimental approaches*, volume xii of *Scientific psychology series.* Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US.

Luce, R. D. (2001). Reduction Invariance and Prelec's Weighting Functions. *Journal of Mathematical Psychology*, 45(1):167–179.

Luce, R. D. and Fishburn, P. C. (1991). Rank- and sign-dependent linear utility models for finite first-order gambles. *Journal of Risk and Uncertainty*, 4(1):29–59.

Luce, R. D. and Fishburn, P. C. (1995). A note on deriving rank-dependent utility using additive joint receipts. *Journal of Risk and Uncertainty*, 11(1):5–16.

Luce, R. D. and Marley, A. a. J. (2000). On Elements of Chance. *Theory and Decision*, 49(2):97–126.

Luce, R. D., Ng, C. T., Marley, A. a. J., and Aczel, J. (2008). Utility of gambling II: risk, paradoxes, and data. *Economic Theory*, 36(2):165–187.

Ludvig, E. A. and Spetch, M. L. (2011). Of black swans and tossed coins: is the description-experience gap in risky choice limited to rare events? *PloS one*, 6(6):e20262.

March, J. G. (1996). Learning to be risk averse. *Psychological review*, 103(2):309.

Marley, A. A. J. and Luce, R. D. (2002). A Simple Axiomatization of Binary Rank-Dependent Utility of Gains (Losses). *Journal of Mathematical Psychology*, 46(1):40–55.

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. pages 55–73.

Miyamoto, J. M. (1988). Generic utility theory: Measurement foundations and applications in multiattribute utility theory. *Journal of Mathematical psychology*, 32(4):357–404.

Moore, M. J. and Viscusi, W. K. (1990). Discounting environmental health risks: New evidence and policy implications. *Journal of Environmental Economics and Management*, 18(2).

Morey, R., Rouder, J., Jamil, T., and Team, R. C. (2015). BayesFactor.

Moscati, I. (2013). How cardinal utility entered economic analysis: 1909-1944. *The European Journal of the History of Economic Thought*, 20(6):906–939.

Newhouse, J. P. (2006). Reconsidering the moral hazard-risk avoidance tradeoff. *Journal of Health Economics*, 25(5):1005–1014.

Phelps, E. S. and Pollak, R. A. (1968). On Second-Best National Saving and Game-Equilibrium Growth. *Review of Economic Studies*, 35(2):185–199.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and others (2007). Linear and nonlinear mixed effects models. *R package version*, 3:57.

Plott, C. R. (1996). Rational individual behavior in markets and social choice processes: the discovered preference hypothesis.

Prelec, D. (1998). The probability weighting function. *Econometrica*, pages 497–527.

Prelec, D. (2004). Decreasing Impatience: A Criterion for Non-stationary Time Preference and Hyperbolic Discounting. *Scandinavian Journal of Economics*, 106(3):511–532.

Prelec, D. and Loewenstein, G. (1998). The red and the black: Mental accounting of savings and debt. *Marketing Science*, 17(1):4–28.

Prelec, D. and Simester, D. (2001). Always leave home without it: A further investigation of the credit-card effect on willingness to pay. *Marketing letters*, 12(1):5–12.

Qiu, J. and Steiger, E.-M. (2010). Understanding the Two Components of Risk Attitudes: An Experimental Analysis. *Management Science*, 57(1):193–199.

Quiggin, J. (1981). Risk perception and risk aversion among Australian farmers. *Australian Journal of Agricultural Economics*, 25(2):160–9.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization*, 3(4):323–343.

Raghubir, P. and Srivastava, J. (2008). Monopoly money: the effect of payment coupling and form on spending behavior. *Journal of Experimental Psychology: Applied*, 14(3):213.

Raiffa, H. and Luce, D. (1957). *Games and decisions*. New York: Wiley.

Rakow, T., Demes, K. A., and Newell, B. R. (2008). Biased samples not mode of presentation: Re-examining the apparent underweighting of rare events in experience-based choice. *Organizational Behavior and Human Decision Processes*, 106(2):168–179.

Rakow, T. and Newell, B. R. (2010). Degrees of uncertainty: An overview and framework for future research on experience-based choice. *Journal of Behavioral Decision Making*, 23(1):1–14.

Read, D. (2001). Is Time-Discounting Hyperbolic or Subadditive? *Journal of Risk and Uncertainty*, 23(1):5–32.

Regenwetter, M., Dana, J., and Davis-Stober, C. P. (2011). Transitivity of preferences. *Psychological Review*, 118(1):42–56.

Regenwetter, M. and Davis-Stober, C. P. (2012). Behavioral variability of choices versus structural inconsistency of preferences. *Psychological Review*, 119(2):408–416.

Rohde, K. I. M. (2010). The hyperbolic factor: A measure of time inconsistency. *Journal of Risk and Uncertainty*, 41(2):125–140.

Ronen, J. (1973). Effects of some probability displays on choices. *Organizational Behavior and Human Performance*, 9(1):1–15.

Rouder, J. N., Morey, R. D., Speckman, P. L., and Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, 56(5):356–374.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., and Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16(2):225–237.

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

Samuelson, P. A. (1937). A note on measurement of utility. *The Review of Economic Studies*, 4(2):155–161.

Sayman, S. and Öncüler, A. (2009). An investigation of time inconsistency. *Management Science*, 55(3):470–482.

Scharff, R. L. (2009). Obesity and Hyperbolic Discounting: Evidence and Implications. *Journal of Consumer Policy*, 32(1):3–21.

Scholten, M. and Read, D. (2006). Discounting by intervals: A generalized model of intertemporal choice. *Management science*, 52(9):1424–1436.

Schunk, D. and Betsch, C. (2006). Explaining heterogeneity in utility functions by individual differences in decision modes. *Journal of Economic Psychology*, 27(3):386–401.

Sidman, M. (1952). A note on functional relations obtained from group data. *Psychological Bulletin*, 49(3):263.

Slovic, P. (1969). Manipulating the attractiveness of a gamble without changing its expected value. *Journal of Experimental Psychology*, 79(1, Pt.1):139–145.

Sneddon, R. and Luce, R. D. (2001). Empirical Comparisons of Bilinear and Nonbilinear Utility Theories. *Organizational Behavior and Human Decision Processes*, 84(1):71–94.

Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of economic literature*, pages 332–382.

Stevens, S. (1936). A scale for the measurement of a psychological magnitude: loudness. *Psychological Review*, 43(5):405–416.

Stiggelbout, A. M., Kiebert, G. M., Kievit, J., Leer, J. W. H., Stoter, G., and Haes, J. C. J. M. D. (1994). Utility Assessment in Cancer Patients Adjustment of Time Tradeoff Scores for the Utility of Life Years and Comparison with Standard Gamble Scores. *Medical Decision Making*, 14(1):82–90.

Stott, H. P. (2006). Cumulative prospect theory's functional menagerie. *Journal of Risk and Uncertainty*, 32(2):101–130.

Strotz, R. H. (1955). Myopia and Inconsistency in Dynamic Utility Maximization. *The Review of Economic Studies*, 23(3):165–180. 02602.

Sutter, M., Kocher, M. G., Rutzler, D., and Trautmann, S. (2013). Impatience and Uncertainty: Experimental Decisions Predict Adolescents' Field Behavior. *American Economic Review*, 103(5):510–31.

Takeuchi, K. (2011). Non-parametric test of time consistency: Present bias and future bias. *Games and Economic Behavior*, 71(2):456–478.

Tanaka, T., Camerer, C. F., and Nguyen, Q. (2010). Risk and Time Preferences: Linking Experimental and Household Survey Data from Vietnam. *The American Economic Review*, 100(1):557–571.

Team, R. C. (2013). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

Tversky, A. and Bar-Hillel, M. (1983). Risk: The long and the short. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9(4):713–717.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323.

Tversky, A., Slovic, P., and Kahneman, D. (1990). The causes of preference reversal. *The American Economic Review*, pages 204–217.

Ungemach, C., Chater, N., and Stewart, N. (2009). Are Probabilities Overweighted or Underweighted When Rare Outcomes Are Experienced (Rarely)? *Psychological Science*, 20(4):473–479.

van de Kuilen, G. (2007). Subjective Probability Weighting and the Discovered Preference Hypothesis. *Theory and Decision*, 67(1):1–22.

van de Kuilen, G. and Wakker, P. P. (2006). Learning in the Allais paradox. *Journal of Risk and Uncertainty*, 33(3):155–164.

van de Kuilen, G. and Wakker, P. P. (2011). The Midweight Method to Measure Attitudes Toward Risk and Ambiguity. *Management Science*, 57(3):582–598.

van der Pol, M. and Cairns, J. (2002). A comparison of the discounted utility model and hyperbolic discounting models in the case of social and private intertemporal preferences for health. *Journal of Economic Behavior & Organization*, 49(1):79–96.

van der Pol, M. and Cairns, J. (2011). Descriptive validity of alternative intertemporal models for health outcomes: an axiomatic test. *Health Economics*, 20(7):770–782.

van Osch, S. M. C., van den Hout, W. B., and Stiggelbout, A. M. (2006). Exploring the reference point in prospect theory: gambles for length of life. *Medical Decision Making*, 26(4):338–346.

Vohs, K. D. and Schooler, J. W. (2008). The value of believing in free will encouraging a belief in determinism increases cheating. *Psychological science*, 19(1):49–54.

von Nitzsch, R. and Weber, M. (1988). Utility function assessment on a micro-computer: An interactive procedure. *Annals of Operations Research*, 16(1):149–160.

von Winterfeldt, D., Chung, N.-K., Duncan, R., and Cho, Y. (1997). Tests of consequence monotonicity in decision making under uncertainty. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(2):406–426.

Wakker, P. and Deneffe, D. (1996). Eliciting von Neumann-Morgenstern Utilities When Probabilities Are Distorted or Unknown. *Management Science*, 42(8):1131–1150.

Wakker, P. P. (2008). Explaining the characteristics of the power (CRRA) utility family. *Health Economics*, 17(12):1329–1344.

Wakker, P. P. (2010). *Prospect Theory: For Risk and Ambiguity*. Cambridge University Press.

Wakker, P. P., Erev, I., and Weber, E. U. (1994). Comonotonic Independence: The Critical Test between Classical and Rank-Dependent Utility Theories. *Journal of Risk and Uncertainty*, 9(3).

Warner, J. T. and Pleeter, S. (2001). The personal discount rate: Evidence from military downsizing programs. *American Economic Review*, pages 33–53.

Wu, G. and Gonzalez, R. (1996). Curvature of the Probability Weighting Function. *Management Science*, 42(12):1676–1690.

Yechiam, E. and Busemeyer, J. R. (2006). The effect of foregone payoffs on underweighting small probability events. *Journal of Behavioral Decision Making*, 19(1):1–16.

Yechiam, E., Rakow, T., and Newell, B. R. (2015). Super-Underweighting of Rare Events with Repeated Descriptive Summaries. *Journal of Behavioral Decision Making*, 28(1):67–75.

Samenvatting|Summary in Dutch

Dit proefschrift onderzoekt de stabiliteit, en adaptiviteit van besluitvorming over tijd en onder risico. Hoofdstuk 2 en 3 behandelen hoe stabiel voorkeuren zijn. Hoofdstuk 2 introduceert een nieuwe methode om geld te verdisconteren over tijd. Hoofdstuk 3 bestudeert reduction invariance, en bevestigt de validiteit van Prelec's compound-invariant kans-gewogen functie. Ten slotte, de laatste drie hoofdstukken onderzoeken hoe adaptief voorkeuren zijn voor verschillende uitkomsten en verschillende context. Hoofdstuk 4 vergelijkt afwijkingen van constante verdiscontering in het geval van gezondheid en geld. Hoofdstuk 5 laat het risicogedrag zien wanneer men moeten kiezen tussen getallen of geld. Hoofstuk 6 onderzoekt dat men compleet verschillende risicovoorkeuren heeft wanneer een besluit gebaseerd is op eigen ervaring, en geeft nieuwe inzichten ten op zichtte van de leemte in de DFD-DFE literatuur.

Hoofdstuk 2 presenteert een directe methode om te verdisconteren waarbij geen kennis over utiliteiten nodig is. Alhoewel alle bestaande methoden om geld te verdisconteren discrete uitkomsten gebruiken hebben de meeste werkelijke besluiten betrekking op stroom uitkomsten, zoals salarisbetalingen, pensioensparen, en hypotheek betalingen. In deze context is het gebruik van de directe methode logischer dan discrete methoden. Echter, onze methode is minder bruikbaar voor beslissingen met één uitkomst.

Hoofdstuk 3 vervolgt het werk van Luce (2001) en test reduction invariance door middel van een experiment. Onze data bevestigt reduction invariance, wat aangeeft dat het gedragsfundament van Prele's compound-invariant kans-gewogen functie empirisch valide is. Een additioneel resultaat van ons onderzoek is dat we tevens de afname van samengestelde weddenschappen hebben onderzocht, wat vaak wordt gezien als een kenmerk van rationele keuze. De meeste proefpersonen (60%) gedroegen zich hierna. De proefpersonen die hiervan afweken, weken voornamelijk af in de richting van equivalenten met hogere zekerheid in samengestelde weddenschappen dan voor de overeenkomstige simpele weddenschappen.

Hoofdstuk 4 evalueert de consequenties wanneer wordt afgeweken van de aanname dat ongeduld constant blijft, en kwantificeert het afnemende ongeduld op het gebied van zowel gezondheid en geld. Het is belangrijk om te weten dat voorkeuren gelijk blijven over tijd voor zowel onderzoek als voor beleidsregels. De resultaten laten zien dat de meeste proefpersonen afweken van de aanname van constant ongeduld en waren meer ongeduldig over tijd voor zowel gezondheid en geld. Proefpersonen weken meer af van constante ongeduld voor gezondheid dan in plaats van geld. Deze domein afhankelijkheid voor verdiscontering laat zien dat tijdsvoorkeuren voor geld weinig validiteit hebben voor gezondheid. Het is daarom niet geschikt voor beleidsmakers om tijdsvoorkeuren in geld te gebruiken voor gezondheid. Tevens, ongeveer één derde van onze proefpersonen toonde zich toenemend ongeduldig, wat duidt dat de bereidwilligheid afneemt over tijd. Deze uitkomsten kunnen niet worden meegenomen in de meeste verdiscontering modellen.

Hoofdstuk 5 beschrijft hoe risicogedrag kan veranderen. Door loterij uitkomsten uit te drukken in geld in plaats van getallen worden mensen minder rationeel: ze waren minder sensitief voor veranderingen in de aannemelijkheidsfunctie. Onze resultaten verklaren tevens de kredietkaart premium: wanneer de waarde van een product onbekend is, zijn klanten met contant geld meer risico-avers dan klanten met een kredietkaart, dit resulteert dat het hebben van een kredietkaart tot meer uitgaven leidt.

Hoofdstuk 6 onderzoekt de DFD-DFE leemte door het oplossen van vier problemen uit vorige studies: steekproeffouten door gelimiteerd onderzoek, ambiguïteit door trekkingen met teruglegging, vertekende geaggregeerde resultaten vanwege het aggregeren van individuele heterogene data, en onduidelijke kans-wegingen vanwege onaangepaste utiliteitsfuncties. Onze resultaten wijzen niet op de algemene uitkomst van een kans-gewogen S-functie in de DFE literatuur: de leemte bestaat nog steeds maar onze resultaten duiden niet naar de directie van een kans-gewogen inverse S-functie. We onderstrepen het belang van leerervaringen om irrationaliteit te verminderen. Besluitvorming gebaseerd op ervaring leidt niet dat een irrationaliteit in een andere irrationaliteit veranderd, maar eerder vermindert irrationaliteit in het geval van goede leerervaring die vrij was van onzuivere metingen en ambiguïteit.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

630 Y. ZENG, *Topics in Trans-boundary River sharing Problems and Economic Theory*
631 M.G. WEBER, *Behavioral Economics and the Public Sector*
632 E. CZIBOR, *Heterogeneity in Response to Incentives: Evidence from Field Data*
633 A. JUODIS, *Essays in Panel Data Modelling*
634 F. ZHOU, *Essays on Mismeasurement and Misallocation on Transition Economies*
635 P. MULLER, *Labor Market Policies and Job Search*
636 N. KETEL, *Empirical Studies in Labor and Education Economics*
637 T.E. YENILMEZ, *Three Essays in International Trade and Development*
638 L.P. DE BRUIJN, *Essays on Forecasting and Latent Values*
639 S. VRIEND, *Profiling, Auditing and Public Policy: Applications in Labor and Health Economics*
640 M.L. ERGUN, *Fat Tails in Financial Markets*
641 T. HOMAR, *Intervention in Systemic Banking Crises*
642 R. LIT, *Time Varying Parameter Models for Discrete Valued Time Series*
643 R.H. KLEIJN, *Essays on Bayesian Model Averaging using Economic Time Series*
644 S. MUNS, *Essays on Systemic Risk*
645 B.M. SADABA, *Essays on the Empirics of International Financial Markets*
646 H. KOC, *Essays on Preventive Care and Health Behaviors*
647 V.V.M. MISHEVA, *The Long Run Effects of a Bad Start*
648 W. LI, *Essays on Empirical Monetary Policy*
649 J.P. HUANG, *Topics on Social and Economic Networks*
650 K.A. RYSZKA, *Resource Extraction and the Green Paradox: Accounting for Political Economy Issues and Climate Policies in a Heterogeneous World*
651 J.R. ZWEERINK, *Retirement Decisions, Job Loss and Mortality*

652 M. K. KAGAN, *Issues in Climate Change Economics: Uncertainty, Renewable Energy Innovation and Fossil Fuel Scarcity*

653 T.V. WANG, *The Rich Domain of Decision Making Explored: The Non-Triviality of the Choosing Process*

654 D.A.R. BONAM, *The Curse of Sovereign Debt and Implications for Fiscal Policy*

655 Z. SHARIF, *Essays on Strategic Communication*

656 B. RAVESTEIJN, *Measuring the Impact of Public Policies on Socioeconomic Disparities in Health*

657 M. KOUDSTAAL, *Common Wisdom versus Facts; How Entrepreneurs Differ in Their Behavioral Traits from Others*

658 N. PETER, *Essays in Empirical Microeconomics*

659 Z. WANG, *People on the Move: Barriers of Culture, Networks, and Language*

660 Z. HUANG, *Decision Making under Uncertainty-An Investigation from Economic and Psychological Perspective*

661 J. CIZEL, *Essays in Credit Risk, Banking, and Financial Regulation*

662 I. MIKOLAJUN, *Empirical Essays in International Economics*

663 J. BAKENS, *Economic Impacts of Immigrants and Ethnic Diversity on Cities*

664 I. BARRA, *Bayesian Analysis of Latent Variable Models in Finance*

665 S. OZTURK, *Price Discovery and Liquidity in the High Frequency World*

666 J. JI, *Three Essays in Empirical Finance*

667 H. SCHMITTDIEL, *Paid to Quit, Cheat, and Confess*

668 A. DIMITROPOULOS, *Low Emission Vehicles: Consumer Demand and Fiscal Policy*

669 G.H. VAN HEUVELEN, *Export Prices, Trade Dynamics and Economic Development*

670 A. RUSECKAITE, *New Flexible Models and Design Construction Algorithms for Mixtures and Binary Dependent Variables*

671 Y. LIU, *Time-varying Correlation and Common Structures in Volatility*

672 S. HE, *Cooperation, Coordination and Competition: Theory and Experiment*

673 C.G.F. VAN DER KWAAK, *The Macroeconomics of Banking*

674 D.H.J. CHEN, *Essays on Collective Funded Pension Schemes*

675 F.J.T. SNIEKERS, *On the Functioning of Markets with Frictions*

676 F. GOMEZ MARTINEZ, *Essays in Experimental Industrial Organization: How Information and Communication affect Market Outcomes*

677 J.A. ATTEY, *Causes and Macroeconomic Consequences of Time Variations in Wage Indexation*

678 T. BOOT, *Macroeconomic Forecasting under Regime Switching, Structural Breaks and High-dimensional Data*

679 I. TIKOUDIS, *Urban Second-best Road Pricing: Spatial General Equilibrium Perspectives*

680 F. FELSO, *Empirical Studies of Consumer and Government Purchase Decisions*