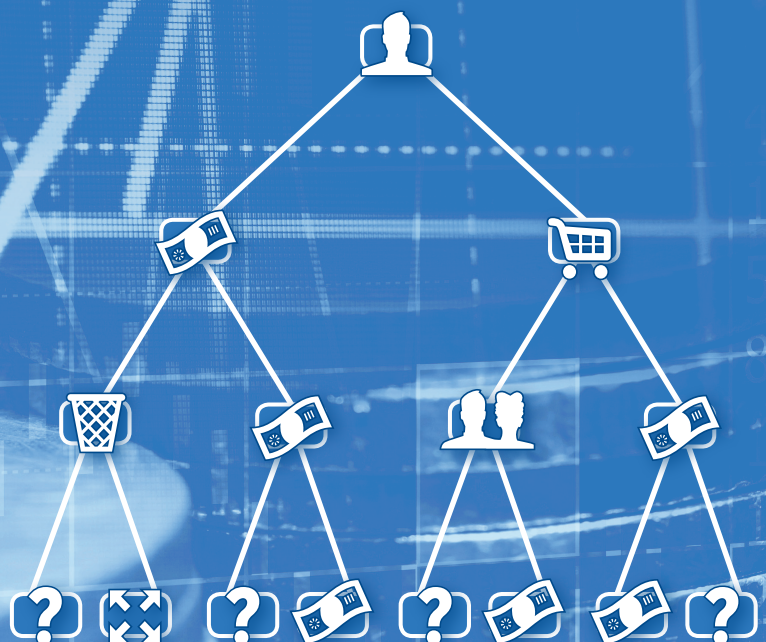


MAXIMILIAN B. TROSSBACH

# Experimental Studies on Risk Preferences, Investment Risk Disclosure, and Motives for Risk-Taking



The dissertation “Experimental Studies on Risk Preferences, Investment Risk Disclosure, and Motives for Risk-Taking” comprises three scientific essays on decision making under risk and uncertainty. Each chapter addresses a specific aspect of the subject matter by presenting experimental evidence on the following research questions:

***Are team decisions as rational as individual decisions?***

***How sensitive is behavior to variations in risk representation when it comes to investing?***

***Is the reduction of effort a prevailing motive when incentive schemes induce risk-taking?***

Maximilian Trossbach has worked on these studies as a doctoral student at the Chair of Empirical Capital Markets Research at WHU – Otto Beisheim School of Management in Vallendar, and as a research fellow at the University of Tilburg and the Leonard N. Stern School of Business at New York University between 2008 and 2014. This book was first published in 2015.

Maximilian Bernhard Troßbach

***Experimental Studies  
on Risk Preferences,  
Investment Risk Disclosure,  
and Motives for Risk-Taking***

Dissertation

For the Attainment of a Doctorate Degree in Economic Sciences  
(Dr. rer. pol.)

at WHU – Otto Beisheim School of Management

Date of Submission: August 28<sup>th</sup>, 2014

First Supervisor: Prof. Dr. Lutz Johanning

Second Supervisor: Prof. Dr. Frauke Freifrau Marschall von Bieberstein





# Acknowledgments

When I was eight years old, I learned knitting at school. Every child in my class was supposed to knit an animal, and I decided to make an elephant. After several weeks of desperate trying, hard work and a lot of help from my mother, my friends and my teacher, I ended up with a little butterfly. At that point in my life, I learned a lot about me – especially that my ambitions often tend to overshoot my capabilities. Ever since then, I set myself targets of which I was confident that I could reach them.

Years later, this dissertation has become my second serious attempt to create something that goes beyond the borders of what I knew to be achievable for me. Since the day I realized this, I have considered this thesis my personal elephant. I would like to express my deepest gratitude to everyone who supported me in this project – without your help, your encouragement and your patience I would not have made it.

First and foremost, I would like to thank my supervisor Lutz Johanning for accepting me as his doctoral student, for his guidance over the past years, for the confidence that he has placed in me, and for many inspiring and enriching conversations.

I am also very obliged to Frauke Marschall von Bieberstein who agreed to be the second supervisor of my doctoral thesis. Her advice in the early stage and especially in the defense of my dissertation has been highly valuable.

I had the pleasure to work together with several great scientists, colleagues and co-authors whose experience in academic work and willingness to discuss and develop ideas has always been an invaluable source of good advice for me. Timur Karabiber, Denis Schweizer and Lars Haß were fantastic colleagues at the Chair of Empirical Capital Market Research at WHU and I am happy to call them my friends. Stefan Trautmann and Gijs van de Kuilen made it possible for me to spend a research semester at the University of Tilburg and to conduct experiments in their laboratory. I am grateful that Yakov Amihud has shared his tremendous knowledge and experience with me in several discussions that had a remarkable impact on my research. He was also my academic sponsor during a research semester at New York University and the opportunity to go there has motivated me more than anything else during my work. I would also like to thank Kay Benk for sharing insights into the psychology of financial decision making with me that awakened my interest in the field. Barbara Luig has always been a lovely and reliable help in virtually all issues during my time at WHU.

In addition to the people mentioned above, I would like to thank many others that contributed to my research through very helpful and motivating comments and discussions. In particular, these are Eva Terberger, Jan Potters, Julian Jamison, Richard Zeckhauser, Jason Shachat, Roman Inderst, Dario De Maio, Lukas Petrikowski, Benedikt Neufang, Marie-Luise Meyer, Alexander Luik, Anna Poimann, Max Müller, Ulrich Hofbaur, Dennis Dlugosch, Marco Hahn and Frank Lohmüller. To name all the other great people, be it colleagues, fellow students, or faculty members who have made my time in Vallendar as exciting and fun as it was would clearly go beyond the scope of this prologue, but the mere thought of so many good memories that I share with you makes me smile!

I am also very grateful to the University of Tilburg and the University of Lancaster who provided funds to conduct my experiments.

Before turning to the most important persons that gave me incredible support throughout my whole life and who are the only ones who could possibly miss on this list of people, I would of course also like to thank the hundreds of participants in my experiments. They form the fundament of this thesis.

My family has been and will be my anchor, my source of joy and self-confidence, and I am thankful for their unconditional support. I have lost faith in myself more than once while I worked on this dissertation and nobody else could have brought me back on track. Their patience, their mental and their active support, countless constructive, but also critical words, and their love have been absolutely decisive for the completion of this thesis. I dedicate my work to my parents Heiner and Sabine and my two sisters Charlotte and Constanze.



# Contents

<b>Acknowledgments</b> .....	3
<b>Contents</b> .....	5
<b>List of Figures</b> .....	8
<b>List of Tables</b> .....	9
<b>1 Introduction</b> .....	12
1.1 Causality and Correlation .....	12
1.2 Advanced Potential for Data Collection in Experiments.....	13
1.3 Summary of the Main Results .....	15
1.4 The Measurement of Risk Preferences .....	17
<b>2 Rational Decisions, Skills, and Group Dynamics under Risk and Ambiguity</b> .....	22
2.1 Introduction .....	22
2.2 Paradoxes in Rational Decision Making .....	28
2.2.1 The Allais Paradox .....	28
2.2.2 The Ellsberg Paradox .....	29
2.2.3 The Zeckhauser Paradox .....	31
2.3 Literature Review on Decision Making Paradoxes .....	32
2.4 Paradoxes in Group Decisions – Experimental Setup .....	34
2.5 Sample .....	38
2.6 Results .....	41
2.6.1 Elicited preferences towards risk and ambiguity .....	41
2.6.2 Individual and Group Rationality in the Allais Paradox .....	51
2.6.3 Individual and Group Rationality in the Ellsberg Paradox .....	53

2.6.4	Individual and Group Rationality in the Zeckhauser Paradox .....	59
2.6.4.1	Russian Roulette Version.....	59
2.6.4.2	Dice Game Version.....	63
2.6.5	Group Rationality and Personal and Social Skills .....	65
2.6.6	Preference Shifts and Personal and Social Skills .....	71
2.7	Discussion.....	76
	Appendix to Chapter 2.....	80

### **3 Risk Disclosure in Key Investor Information Documents, Risk Perception, and Investment Decisions..... 100**

3.1	Introduction .....	100
3.2	Experimental Design .....	103
3.2.1	Reference Study from the US market .....	103
3.2.2	Selection of Funds .....	106
3.2.3	Experimental Stages .....	107
3.2.4	Investment Stage .....	108
3.2.5	Self-Assessment Stage .....	110
3.2.6	KID Assessment Stage .....	111
3.3	Results .....	111
3.3.1	Sample Statistics .....	111
3.3.2	Comparison of Fund Characteristics over Experimental Treatments .....	115
3.3.3	Information Sets and Risk Perception .....	117
3.3.4	Portfolio Selection and Willingness to Invest .....	124
3.3.5	Relevance of Fund Characteristics and Suggestions for Improvement of KIDs 136	
3.4	Discussion.....	139
	Appendix to Chapter 3.....	142



<b>4</b>	<b>Making Easy Money – Do We Trade Risk for Laziness?</b>	170
4.1	Introduction	170
4.2	Experimental Design	172
4.3	Results	175
4.3.1	Description of the Samples	175
4.3.2	Effort Exertion and Lottery Choices	176
4.3.3	Risk Preferences and Lottery Choices	183
4.3.4	Beliefs	185
4.4	Discussion	186
	Appendix to Chapter 4	189
<b>5</b>	<b>Final Remarks</b>	201
	<b>Bibliography</b>	203

# List of Figures

Figure 2.1: Risk Preference Parameters of Linear Utility Functions .....	45
Figure 2.2: Hypothetical Group Risk Preference and Risk Shift .....	47
Figure 2.3: Hypothetical vs. Actual Differences in the Willingness to Pay (Zeckhauser Paradox, Russian Roulette version) .....	62
Figure 3.1: Synthetic Risk and Reward Indicator used in KIDs .....	101
Figure 3.2: MCC with Additional Rating.....	104
Figure 3.3: Graphical Display of Annual Returns in MCC.....	105
Figure 5.1: Financial Assets of Private Households in Germany 2013 in Billion Euros .....	201

# List of Tables

Table 2.1: Acceptance of EUT and SEUT Axioms .....	39
Table 2.2: Peer Evaluations, Final Grade from Secondary School, and Assessment Center Scores .....	40
Table 2.3: Risk and Ambiguity Preferences for Individuals and Groups .....	42
Table 2.4: Categorized Risk Preferences for Individuals and Groups .....	42
Table 2.5: Categorized Ambiguity Preferences for Individuals and Groups .....	43
Table 2.6: Probability Conversion Factors and Subjective Probabilities implied by Individual and Actual Group Decisions .....	49
Table 2.7: Individual Majorities and Group Decisions on Group Level in the Allais Paradox (common consequences, N = 36) .....	51
Table 2.8: Individual Majorities and Group Decisions on Group Level in the Allais Paradox (common ratio, N = 36) .....	52
Table 2.9: Consistency with SEUT for Majorities of Individuals and Group Decisions on Group Level in the Ellsberg Paradox (N = 36) .....	54
Table 2.10: Consistency with SEUT of Individual and Common Group Decisions on Individual Level in the Ellsberg Paradox (N = 108) .....	54
Table 2.11: Consistency with Elicited Ambiguity Preferences (measured by $\gamma_i$ and $\gamma_j$ over all gambles) for Majorities of Individuals and Group Decisions on Group Level in the Ellsberg Paradox (N = 36) .....	56
Table 2.12: Consistency with Elicited Ambiguity Preferences (measured by $\gamma_i$ and $\gamma_j$ over all gambles) of Individual and Common Group Decisions on Individual Level in the Ellsberg Paradox (N = 108) .....	56
Table 2.13: Consistency with Elicited Ambiguity Preferences (measured by $\gamma_i$ and $\gamma_j$ over gambles with $p^{obj} = 0.5$ ) for Majorities of Individuals and Group Decisions on Group Level in the Ellsberg Paradox (N = 36) .....	57

Table 2.14: Consistency with Elicited Ambiguity Preferences (measured by $\gamma_i$ and $\gamma_j$ over gambles with $p^{obj} = 0.5$ ) of Individual and Common Group Decisions on Individual Level in the Ellsberg Paradox (N = 108).....	57
Table 2.15: Individual Majorities and Group Decisions on Group Level in the Zeckhauser Paradox (Russian Roulette, N = 36).....	60
Table 2.16: Consistency of Individual and Common Group Decisions on Individual Level in the Zeckhauser Paradox (Russian Roulette, N = 108) .....	60
Table 2.17: Individual Majorities and Group Decisions on Group Level in the Zeckhauser Paradox (Dice Game, N = 36).....	63
Table 2.18: Consistency of Individual and Common Group Decisions on Individual Level in the Zeckhauser Paradox (Dice Game, N = 108).....	64
Table 2.19: Determinants of Consistency of Group Decisions in the Ellsberg Paradox (N = 36) .....	67
Table 2.20: Determinants of Consistency of Group Decisions in the Dice Game version of the Zeckhauser Paradox (N = 36).....	69
Table 2.21: Personal Skills and Individual Risk Preference Shifts.....	73
Table 2.22: Personal Skills and Individual Ambiguity Preference Shifts.....	75
Table 3.1: Educational Level and Professional or Educational Backgrounds in Financial Investments.....	112
Table 3.2: Self-categorization in Financial Literacy .....	113
Table 3.3: Self-categorization in Risk Tolerance in Financial Investments.....	113
Table 3.4: Financial Literacy and Risk Tolerance by Professional / Educational Background and Gender .....	114
Table 3.5: Age, Financial Literacy, and Risk Tolerance by Treatment .....	115
Table 3.6: Usability of KIDs for Comparison of Fund Characteristics .....	116
Table 3.7: Assigned Riskiness Rankings per Treatment.....	118
Table 3.8: Correlation between Assigned Fund Risk and SRRI Values .....	120
Table 3.9: TOBIT Model (Interaction Effects) Risk Estimation.....	123



Table 3.10: Allocations per Treatment.....	125
Table 3.11: Allocations to Checking Account .....	127
Table 3.12: Allocations to Fund BLUE .....	129
Table 3.13: Allocations to Fund GREEN.....	131
Table 3.14: Allocations to Fund RED .....	132
Table 3.15: Allocations to Fund WHITE .....	133
Table 3.16: Value Weighted SRRI of Allocation .....	135
Table 3.17: Importance of Fund Characteristics per Treatment.....	137
Table 4.1: Demographics .....	176
Table 4.2: Effort Levels by Stage and Experimental Group .....	177
Table 4.3: Pooled Regression Effort Levels Stages 3, 4, 5 .....	178
Table 4.4: Probit Regression Lottery Choice with endogenous regressor .....	180
Table 4.5: Effort Levels by Stages and Frequency of Lottery Choices .....	182
Table 4.6: Value Constellations between Certainty Equivalents and Earnings from Slider Task versus Choices of Payment Mode .....	184

# 1 Introduction

This dissertation consists of three articles on experiments in the field of (financial) decision making when risk or uncertainty is involved. Before we turn to the outlook on the actual content of this thesis, it is worth pointing out two distinctive features of experiments as a research method. First, experiments offer the opportunity to detect causal relationships between economic factors. Second, experiments are feasible to obtain data that is not easily accessible in real settings.

Following these general remarks, section 1.3 summarizes the experiments and the main findings that are presented in the main body of this dissertation in chapters 2, 3, and 4. Section 1.4 addresses some important methodological issues concerning the measurement and description of risk and risk preferences and compares the different approaches that are used in the three experiments in this thesis.

## 1.1 Causality and Correlation

An experiment can be considered a controlled process to generate primary empirical data that is tailor-made to answer a specific research question. Thereby, the central element is the experimental variation. The idea is to create at least two environments in which – in the social sciences – the behavior of the participating test persons is observed. Ideally, the environments, often called the control group and the experimental condition(s) or treatment(s), are identical except for one single aspect that is exogenously changed by the experimenter. If there is any difference in the observed behavior of the subjects in two environments of the same experiment, this difference is *caused* by the experimental variation. The power to detect causal relationships is the key advantage of experiments over many other empirical research designs, and it is useful to understand how this can be achieved.

The majority of empirical studies in the business and economics literature are correlational, i.e. they identify relationships between various economic factors that enhance predictions of unknown factors if others are known. Looking, for example, at regression analysis as the standard statistical method to identify a functional relationship between two variables, observing a correlation between two variables  $X$  and  $Y$  means that values of one variable (the dependent variable  $Y$ ) can be predicted more precisely if the other variable (the independent variable  $X$ ) is incorporated in the prediction model of the form  $y_i = a + bx_i$  with  $a$  and  $b$  being the OLS estimators of the regression model, compared to simply using the sample mean

as the predictor, i.e.  $y_i = \bar{y}$ . However, even though the denotation as “independent” and “dependent” variables suggests that variation in  $Y$  is actually caused by the variation in  $X$ , this must not be inferred from regression results, unless a critical assumption holds. The correlation between two variables is symmetric by construction, which means that a reversed model  $x_i = a + by_i$  yields equally good predictions. The variation in  $X$  might therefore as well be caused by the variation in  $Y$ , or there might also be reciprocal causations. Only if we know that  $X$  cannot be caused by  $Y$ , it is valid to interpret the observed correlation between the two variables as a causal relationship of  $X$  on  $Y$ . A reverse causation of  $Y$  on  $X$  can be excluded if the variation in  $X$  is exogenous, which is, for instance, the case when it is controlled by an experimenter. If  $X$  is a dummy variable that indicates whether participant  $i$  takes part in the control group or the experimental group, and the participants are assigned randomly to these two conditions, differences in behavior (measured as  $Y$ ) are caused by the experimental variation if the resulting parameter  $b$  from the regression model  $y_i = a + bx_i$  is significantly different from 0. A reverse causation can be excluded, because the test persons can neither self-select into one of the experimental conditions, nor does the assignment to the groups depend in any way on behavioral differences between the subjects – which are not even known at the time when the random assignment is made. Thus, if the experimental condition differs from the control group in only one aspect, it is possible to infer that differences in behavior are caused by the variation.

## 1.2 Advanced Potential for Data Collection in Experiments

To have full control over the experimental environment(s) also means that the types of data that can be collected in an experiment are manifold. When working with field data, the researcher needs to rely on observations of actual choices and disclosed information, but it is difficult if not impossible to control for mental constructs like preference systems that lead to the actual decisions or hidden actions that lead to the disclosed results. Furthermore, in order to address certain research questions, field data might simply not be available in the form of secondary data and collecting new data might be difficult due to legal, ethical, or other restrictions. In these cases, hypothetical experiments are a loophole to overcome these problems.

Learning about individual preferences is an integral pre-condition if one wants to test theories on decision making, or strives to evaluate the consistency and rationality of decisions. The second chapter of this thesis addresses these issues elaborately, but also in the other chapters,

the measurement of risk preferences plays an important role. I will discuss different methods to elicit preferences that have been applied in the three studies presented in this dissertation in section 1.4, and of course in the respective chapters where the experimental setups are described.

The third chapter of this thesis is based on a controlled hypothetical choice experiment where the participants can invest in different financial products. The collection of field data would have imposed a conflict with the existing regulation of the financial products that are involved and financial advisors would have been at risk to be accused for miscounseling.

The experiment that is presented in the fourth chapter addresses a typical hidden action problem. In the field, i.e. in real-world data, what we see are usually outcomes like for example the realized profit of a company or the resulting market price for a traded good. However, it is often unclear what kind of behavior of the involved agents has actually led to the observed outcomes. Classical issues are the separation of luck and skills, performance and risk-taking, or contribution and free-riding. All this can be made completely transparent in feasible experimental designs. Furthermore, the experimenter can control the degree of transparency that is faced by each group of subjects and also by herself to allow for more or less opportunities to commit hidden action, and to discriminate the degrees of privacy towards other participants in the experiment and the experimenter. In the experiment presented in the fourth chapter, we design two payoff-generating processes that are strictly independent from each other. We unambiguously observe for which of the two alternatives each test person decides, even though both can produce identical payoffs. In reality, where only payoffs are observable and verifiable, this would be difficult to identify.

To collect, to analyze and to publish personal and potentially sensitive data requests a great deal of responsibility from any experimenter. It is crucial that the participants of an experiment can rightfully trust that their data is treated confidentially in a way that it will be anonymized and that the reader of an experimental study cannot infer any personal information from the published results. Furthermore, it is necessary that the test persons actually know that they are taking part in an experiment, that they know the general purpose of the experiment, that they understand their participation is voluntary, and that they receive an appropriate briefing and de-briefing from the experimenter before and right after the experiment. This includes the provision of the experimenter's contact details to the participants so that they can ask questions at any time. In order to receive funds for the



experiments presented in this dissertation, the research proposals and experimental designs have been examined and approved by ethics committees.

### **1.3 Summary of the Main Results**

In chapter 2 we analyze individual and group decisions under risk and ambiguity with a special focus on preference shifts and rationality. Risk is defined as an uncertain prospect with known outcomes and known probabilities for these outcomes, while the probabilities in ambiguous prospects are unknown. We find that groups are not more rational than individuals – depending on the criteria we use to classify choices as rational or irrational, they can even perform significantly worse. We find a general tendency of groups being more risk neutral than expected, while there is no overall preference shift in ambiguous decisions. By merging the experimental data with the results from an assessment center and from peer evaluations, we find that rationality of group decisions and preference shifts caused by group formation can partly be explained by personal and social skills, but individual rationality cannot.

Following this rather theory-driven and abstract – risky and ambiguous prospects are predominantly modeled as simple binary lotteries – examination of the interplay between preferences and decisions, the next chapter turns to more complex, but also more practical, investment decisions. Investment risks are generally not describable by precise probabilities for a small number of potential outcomes, as it is done in chapter 2, and different risk factors come into play. In practice, this leads immediately and inevitably to the question how these risks should be disclosed to potential investors in a way that they are well comprehensible and comparable.

In chapter 3 we vary the descriptions of the risk and reward profile of selected investment funds in Key Investor Information Documents (KIDs) and analyze the effects on risk perception and investment decisions in an experimental study. KIDs are highly standardized product information sheets that are required by financial regulators in the European Union for investment funds. The risk and reward profile usually consists of a numerical scale and a verbal description. We find that, as expected, removing both elements together reduces the ability to estimate the risk of the available investment options correctly. However, if only a verbal description is provided, risk estimations are even less consistent than in the case without any form of risk disclosure. As a consequence, portfolio risk increases considerably. Thus, the purely verbal explanation is confusing and the scale is important to interpret non-quantitative verbal information correctly. Therefore, we argue that information on a fund

property that is presented verbally should be accompanied by a meaningful numerical scale whenever possible. We also discuss potential extensions of KIDs and the application of a numerical risk scale to other financial products.

Capital investments can be interpreted as the willingness to take on financial risks – and potentially to accept other negative aspects like limited availability and fungibility of the invested capital, i.e. illiquidity – in anticipation of a positive return. In this sense, capital is an income-generating factor. Is there a similar risk and return trade-off when income is not, like in classical investment decisions, a function of capital, but regular labor income that depends on the performance in a given task? Can we observe the willingness to take on risks in order to receive an equally high income with no or lesser input of working effort?

In chapter 4 we study the trade-off between risk-taking and performance as alternatives to generate income. In a laboratory experiment, we study how the possibility to choose an effort-free, risky outside option influences effort provision in a computerized task under a convex payoff scheme that yields highly attractive payments for high effort levels, but relatively low payments for low effort levels. Especially in the financial industry, such compensation schemes have been employed in the form of bonus payments. Intended as a measure to incentivize high effort, such payment schemes are often found to induce risk-taking as an unwanted side effect. Exerting a high effort in a limited time is difficult on the one hand, but the resulting payoffs are risk-free. Choosing a lottery instead as the payoff generating process introduces risk, but it is possible to earn money totally effort-free. Of course, skills and risk preferences vary between subjects. Thus, neither of the two payment options is *ex ante* preferable to everybody. We find that risk-taking is very common when the outside option is available, but effort provision is not significantly influenced by this. Instead of making *ex ante* decisions in favor of the risky option and consequentially reducing effort, we observe that the broad majority of the test persons exerts effort and decides about the preferred payment mode *ex post*. The *ex post* choices are only consistent with previously elicited risk preferences when subjects should rationally choose the payment from the effort task, but not when they actually should make use of the risk-taking option.

In chapter 5, some final remarks conclude this dissertation.

## 1.4 The Measurement of Risk Preferences

All three experiments that are presented in this dissertation address different aspects of decision making when uncertainty in the form of risk or ambiguity is involved. The test persons need to decide between two or more options that are characterized by different degrees of riskiness or they are asked to evaluate different risky prospects in monetary terms. When dealing with decisions in an uncertain environment in a controlled experiment, three factors that are closely interrelated to each other are in the focus of the research design: the measurement of risk, the description of risk in the experiment, and the measurement of the test persons' risk preferences. Measuring risk is any process that is feasible to create a rank order of different prospects according to their riskiness by an objective criterion. This is done by the experimenter during the design of the experiment and plays an important role in the subsequent analysis in order to describe the available choice options and risk determinants as precisely as possible to the reader of the experimental study. These criteria can be probabilities for potential scenarios, moments of a density function of a stochastic outcome variable, or other measures like Value at Risk or volatility. As a description of risky prospects to the participants of an experiment, one could simply disclose the applied risk measures to them, but depending on the intention of the research design, it might be reasonable to choose different ways to describe risk. First, it is likely that condensed numerical risk measures are difficult to understand so that other descriptions may be more informative to the test persons. Second, the description of risk itself may link to the experimental variation. This is for example the case when risky and ambiguous prospects shall be compared, or different methods for risk disclosure are tested against each other. Finally, the measurement of risk preferences is the attempt to collect individually specific data from each test person that enables the researcher to infer the value systems or motives that determine the choices made by the subjects. This is surely the most problematic factor of any experimental analysis of choice behavior, as the modelling of preference systems usually involves a lot of explicit and implicit theoretical assumptions and preconditions in order to make it a tractable issue. For example, a utility function – the mere existence of such a function is a theoretical assumption – usually has very few parameters, while it is known that the number of both, conscious and subconscious influential factors in the perception and evaluation of risk, can be large.<sup>1</sup> On the other hand, choice options in an experimental environment are often described in a very

---

<sup>1</sup> Breakwell (2007) gives an excellent overview on psychological insights about risk behavior.

simple manner and it may therefore be fair to assume that choices in such a controlled environment are actually governed by a rather rudimentary preference system. The researcher should choose a method to measure risk preferences that fits to the actual choices that the test persons have to make during the experiment, and this largely depends on the way risky prospects are described to them.

In the following, we will discuss the different approaches to risk descriptions and risk preference measurements that are used throughout this dissertation. Two of the experiments, namely the one on individual and group rationality and the one on risk and real effort, are designed in a rather abstract manner with regards to the representation of risk, while the experiment on KIDs and investment decisions strives to demonstrate the situation as realistic as possible.

In the chapter on risk and real effort, the modelling of risk in the experimental design is surely the simplest among the three studies. It is described as a lottery with two outcomes and the probabilities of occurrence for each outcome. The lottery is never changed throughout the entire experiment. In such a case, it is common practice to elicit each test person's certainty equivalent for the lottery in order to determine her risk preference. A certainty equivalent is a monetary valuation of a lottery. The idea is to ask a test person how high a fixed payment needs to be to leave her indifferent between receiving this fixed amount and playing the lottery. In our experiment, we do this by using a choice list with binary choices between the lottery and ascending fixed payments. For each pair, the test person is asked whether she prefers the lottery or the fixed payment, so that we can infer the certainty equivalent from the answers. Using such a list has one clear advantage over simply asking the subjects to write down their certainty equivalent: they are forced to make several decisions and therefore to reflect whether they would still accept the fixed payment if it were a little lower, or if the lottery would still be more attractive if the fixed payment increased a little bit. The results are therefore very likely to be accurate, and the process is still not very time-consuming. However, choice lists do not necessarily produce consistent answers. Some participants in the pretests that we performed when we designed the experiment seemed to make their choices randomly so that, for example, they preferred the lottery when the fixed payment was 1 Euro, then they preferred the fixed payment when it raised to 2 Euros, but then they chose to play the lottery again when the fixed payment was 3 Euros. We therefore needed to enforce the



participants to switch over from the lottery to the fixed payment exactly once when going down the list in order to sustain transitivity of the elicited preferences.<sup>2</sup>

In the chapter on individual and group rationality, there are several different decisions that we analyze with regards to their consistency with the elicited preferences. They are characterized by a small (three in the Allais Paradox, two in all others) number of monetary outcomes that are achievable and by the probability of occurrence for each outcome.<sup>3</sup> Hence, the risks that need to be evaluated in the decision tasks are diverse. Therefore, we decided to elicit the test persons' risk (and ambiguity) preferences under a broad selection of lotteries that cover the bandwidth of potential outcomes in the actual choice tasks and to deduct one parameter that describes an individual's (or a group's) risk or ambiguity preference. The test persons need to state a selling price for each lottery that leaves them indifferent between selling the lottery at this price and playing it. The mechanism incentivizes the subjects to state their selling prices in a way that they reflect their actual certainty equivalent. The technical details are described at full length in section 2.6.1. To install a choice list for each lottery would be too time-consuming for the participants to work through. In fact, stating the selling prices directly for 24 lotteries already lasts very long, and this is indeed the major drawback of this procedure. In return, the test persons have the opportunity to reflect their risk preferences thoroughly and the procedure is therefore likely to produce plausible results. In fact, we observe an unusual high degree of consistency in the Allais Paradox in our experiment which might be due to the fact that the test persons do not answer the questions spontaneously but they are trained to evaluate risky prospects when they turn to the task.

In the KID experiment the way risk is displayed to the participants is much more complex than in the other two studies. The test persons need to choose a portfolio of financial products with different risk and reward profiles that cannot be described appropriately through expected values and probabilities for certain scenarios. Instead we follow a new classification model for the riskiness of investment funds that is based on historical volatilities and a qualitative discussion of relevant risk factors. This immediately gives rise to the problem that the risk preferences of the test persons cannot easily be converted into the taxonomy of the classification model like it can be done with certainty equivalents and binary lotteries. Therefore, we apply a very easy self-categorization task to elicit risk preferences in this

---

<sup>2</sup> Section 2.1 provides an overview on the underlying axioms of rational decision making.

<sup>3</sup> Obviously, the probabilities are partly unknown in the Ellsberg Paradox. The measurement of ambiguity preferences is still conducted analogously to the measurement of risk preferences, so we can ignore this for the moment.

experiment. In order for a risk classification model to be useful in an investment decision, the investor needs to learn to translate her preferences, motives and attitudes into the “language” of the model and then to decide which alternative fits best to her.

Overall, the experimenter needs to find a method for the elicitation of risk preferences that is appropriate for the situation that shall be analyzed and needs to put special attention to simplicity (to prevent mistakes caused by misunderstandings), time constraints (to keep the test persons motivated to finish the experiment), and incentive compatibility (to make sure the test persons answer truthfully). Incentive compatibility is especially important when decisions are hypothetical, like in most laboratory and questionnaire-based experiments. To overcome this problem, some research designs link the elicitation of risk preferences to actual behavior; for example Lejuez et al. (2002) introduce the “balloon task” in which the test persons can decide each time to blow up a balloon a little more or not. If they do it, money is added to their account, but each time there is a chance that the balloon explodes and all gains are lost. Anderson and Mellor (2008) even link risk aversion to real life behavior out of the experimental environment like smoking, heavy drinking, overweight and seat-belt usage.

If the experimenter does not observe “real life” decisions or behavior, the test persons are usually incentivized to answer questions truthfully and to decide realistically through the payment mechanisms in the experiment. In the experiment on individual and group rationality the payoff that the participants can earn depends completely on the outcomes of the lotteries in the experiment. In the real effort experiment, the total payoff consists of a fixed show-up fee and a variable effort- and lottery-dependent part. In the KID experiment, the test persons are rewarded for their participation, but the payoff does not depend on their choices and answers. Each payment system, fixed and variable, has certain strengths and weaknesses, and once more, it is the challenge of the experimenter to choose a payoff scheme that best fits the requirements of the research agenda. Fixed payments in the form of show-up fees secure that the participants are not disappointed in case they are unlucky. Satisfied test persons are more likely to return when being recruited for another experiment and they also motivate others to take part. On the other hand, there is the danger that some people might want to earn the show up fee without really paying attention to the questions and tasks in the experiment which impairs the data quality. Incentivizing the subjects to answer truthfully through variable and choice-dependent payment schemes can overcome this problem, but they need to be carefully designed in order to produce the desired effects. For example, a researcher might be interested in the subjects’ behavior in a specific, isolated situation, but throughout the experiment,

several variations are tested so that, in effect, each subject makes a large number of decisions within one experimental session – for example whether she prefers to play a binary lottery or to receive a fixed payment. If all decisions were relevant for the payoff in the sense that all lotteries would be played out and the sum of the gains were paid to the subject, she will in fact receive a payoff that is equal or close to the sum of the expected values of the lotteries. However, if each lottery is considered separately, she will actually never receive the expected value, but only one or the other outcome. If the subject anticipates this, it is likely that her choice behavior resembles risk neutrality, even though she would be risk averse in a one-shot decision. Therefore, in the two experiments where we use variable payment schemes, a small number of choices are determined randomly after the completion of the experiment so that the test person does not know which ones will determine her payoff.

In the KID experiment we refrain from a variable payment system. The way risk preferences are measured is not by the observation of choices, but through self-categorization for a broad class of choices, namely investment decisions in general. Also the portfolio selection that is made by the test persons is purely hypothetical and it is difficult to reward their choice of a certain combination of investment funds variably, and especially in a way that actually reflects the risk and reward profile of the resulting portfolio and that is easily comprehensible and transparent for the test persons. As we analyze the effects of different ways to describe the risk and reward profiles of the (hypothetically) available investment options in this experiment, there would actually be the danger that the test persons' choice behavior would be fully governed by an installed variable payment scheme and the way it is explained to the subjects rather than by the provided information on the available choices. Furthermore, many questions in the KID experiment are posed in order to learn about the participants' opinions, experiences, and personal backgrounds, and it is generally not desirable to influence the answers through incentives. The main purpose of a payment scheme in this experiment is to generate a large number of complete observations (i.e. to have many people working their way through the questionnaire until the end), and we achieved this by announcing a raffle of vouchers for a popular online shop among those participants who submit a code word that is disclosed on the very last page of the survey.

Finally, the way participants are reimbursed for their participation in experiments is also partly determined by the culture of the laboratory and the expectations of the potential subject pool. For example, at Tilburg University, the payment of a show-up fee is simply a standard that should be met by any experiment on campus.

## 2 Rational Decisions, Skills, and Group Dynamics under Risk and Ambiguity<sup>4</sup>

### 2.1 Introduction

In this chapter, we report the results of an extensive choice experiment that investigates the role of personal and social skills in both, individual and team decisions. Using three well-known tasks, namely the Allais Paradox, the Zeckhauser Paradox, and the Ellsberg Paradox, we test whether teams can decide more rationally than single persons and if and how skills affect rationality. In order to categorize decisions as rational or irrational, it is necessary to have a normative theoretical benchmark model. For decisions under risk, as in the Allais Paradox and the Zeckhauser Paradox, we use Expected Utility Theory (EUT, see von Neumann and Morgenstern (1944)). For decisions under ambiguity in the Ellsberg Paradox, we use Subjective Expected Utility Theory (SEUT, introduced by Savage (1954)), which assumes neutrality towards ambiguity, and elicited ambiguity preferences in order to test consistency of choices with preferences. The terms rationality and consistency are therefore used interchangeably throughout this chapter.

In many realistic settings, professional and private, decisions are made by teams. A common explanation for this is of course the notion that teams minimize error, benefit from the bundling of competence, experience, and other factors. Thus, they are said to assure rationality in a wider sense. However, we know that individuals do not act rational all the time and it is not clear how individual preferences and competences aggregate to a (potentially less flawed) team decision. First, a team decision is likely to be determined by predominant existing preferences, i.e. by a majority. However, single individuals might also be able to convince others to change their preferences. Second, there are also group dynamic processes in preference formation that need to be considered when explaining decision making in groups. In the literature, we find experimental evidence for group polarization effects towards both, more risky and more cautious decisions. Stoner (1968) was the first to explain such polarization by values that are widely held in the community, with “risky shifts” occurring when the value system tends to support the riskier of two options, and “cautious shifts” vice

---

<sup>4</sup> The original essay, entitled “Rational Decisions, Skills, and Group Dynamics under Risk and Ambiguity”, is joint research work of Lars Helge Haß, Denis Schweizer, and me. See Haß et al. (2013), unpublished manuscript.

versa. Furthermore, when individuals in a group perceive themselves particularly more risk-loving (risk-averse) than other people in the community, this group's decision reflects a "risky shift" ("cautious shift"). Bem et al. (1964) explain the tendency towards more risk in groups by a weaker responsibility that each group member feels for the consequences of the decision. Schlenker and Weigold (1991) investigate shifts in risk preferences of individuals when the framing of the choice task is changed and find that risk averters who are held accountable decide more cautiously than those who are not. In a recent study, Vieider (2011) analyzes interactions between financial incentives and incentives from accountability considerations. Even though financial incentives seem to overrule effects from accountability in risky decision making concerning losses, accountability effects are found to play a major role in the reduction of preference reversals between different frames and when trading off simple and compound lotteries. While the overall direction is unclear ex ante, all the above-mentioned effects are likely to induce shifts of a group's revealed risk preference compared to the average of all its members.

Other studies test for group polarization effects in investment decisions, also reporting ambiguous results. Bär et al. (2011) find that management teams of US mutual equity funds follow more diversified and more balanced investment strategies than single managers, and attribute this to the descriptive superiority of the diversification of opinions theory over group polarization. Barber et al. (2003) on the other hand show that stock clubs are more likely than individual investors to favor stocks for which there is a "good reason" available to buy them. Obviously, convincing or at least appealing reasons foster group polarization, as they facilitate coordination between various options.

We extend the existing literature<sup>5</sup> on the rationality of group decisions and preference formation by controlling for a number of personal and social skills of the test persons. We find that individual rationality is not determined by these skills, but group rationality can partly be explained by the skills of the team members. Interestingly, even though all of the attributes that we use in our analyses have a clearly positive connotation in the light of decision making and teamwork, some of them also have a significant negative impact on rationality. Furthermore, we see structural differences in several skill variables when we compare individual ambiguity preference shifts that are caused by group formation. For risk preference shifts, skills seem to play only a subordinate role. We conclude that the

---

<sup>5</sup> See section 2.3 for a separate review of the existing literature on group decisions in the paradoxes that we analyze.

investigation of the role of skills and potentially other psychological variables in social choice is a fruitful field for further research, as there are several issues that we cannot answer within the framework of our experiment. For example, it is unclear how a skill level of a team is appropriately measured. We present different approaches, but the question which one is the best has to remain open.

Note that, despite the analysis of group decisions, our framework is not game theoretical. In game theory, the payoffs of any single player depend on the actions of other players that are non-controllable – just observable or predictable in some cases. In classical decision theory the determinants of a decision, i.e. one's own preferences and expected payoffs conditional on actions, are fully transparent. In order to consider our experiments purely decision theoretically, it is helpful to envisage the decision of the group as made by an individual that acts as a representative. Beforehand, the group needs to agree on a common position by convincing each other through argumentation. Note that we leave the structure of the decision making process up to the groups. With fixed rules for group decision making, especially majority voting, irrational preference orders can prevail in spite of consistent individual preferences of each group member, caused only by the design of the voting scheme. This phenomenon has been demonstrated early in the well-known Condorcet Paradox (Condorcet, 1785). Arrow (1963) even shows in his impossibility theorem, that for certain axioms imposed on the decision making process of a group, it is not possible to translate the group's individual preferences into a coherent collective decision. It can be shown that the Condorcet Paradox is a special case of Arrow's impossibility theorem. However, the theorem only addresses decisions which are beyond the scope of our experiment, i.e. with at least three alternatives to choose from. Furthermore, it is always possible to come to a decision by installing dictatorship, which is ruled out by the axioms set up by Arrow. Still, it is conservative not to impose any rules on our experimental groups. For the special case of ambiguous choices see Hylland and Zeckhauser (1979). They derive theoretically that unanimous (non-dictatorial) choices are impossible to achieve when prior beliefs on probabilities and utilities are aggregated separately. We do not require such separation from our test persons.

Before we introduce the three paradoxes that we analyze, we provide a brief overview on the foundations of decision theory which are relevant for this article, namely EUT and SEUT. Under the assumption that these theories actually *describe* individual and group preferences, choices in the aforementioned paradoxes can be classified as rational or irrational. The

prerequisite for these theories to have any *normative* power is the acceptance of a set of underlying axioms regarding the structure of individual preferences on consumption or wealth under uncertainty. It is thereby important to distinguish two kinds of uncertainty in the framework of decision theory: “risk” defines situations of uncertain consequences or states of nature, but with objectively known probabilities assigned to each consequence or state. “Ambiguity” defines situations where the probabilities of uncertain consequences or states of nature are unknown. Depending on the circumstances, a slightly different set of axioms applies.

- 1) The completeness axiom
- 2) The transitivity axiom
- 3) The continuity axiom
- 4a) The independence axiom (under risk)
- 4b) The sure-thing principle (under ambiguity)

Under risk, and applying the independence axiom 4a as postulated by Samuelson (1952), rational behavior can be described by EUT, while under ambiguity axiom 4b, the sure-thing principle, applies instead and rational behavior can be described by SEUT. All other axioms apply for each decision framework. Axioms 1 and 2 ensure that the whole range of consumption goods is covered by a consistent preference order, i.e. it holds either  $X \succeq Y$  or  $X \preceq Y$  for every randomly chosen pair of consumption goods  $X$  and  $Y$ <sup>6</sup> (completeness) and for all consumption goods  $X$ ,  $Y$ , and  $Z$  it holds true that if  $X \succeq Y$  and  $Y \succeq Z$ , then  $X \succeq Z$  (transitivity). Axioms 1 and 2 ensure rationality and stringency of any preference structure regardless of the specific decision rule that is applied to it. Preferences that satisfy axioms 1 and 2 can be described by an ordinal utility function, i.e. any monotone transformation of the function yields the same preference order in the sense that the most preferred option is always the one with the highest utility value. Utility is a theoretical construct that translates wealth or consumption into a level of “well-being” that is determined by the individual’s utility function.

---

<sup>6</sup> Note that consumption goods  $X$  and  $Y$  do not necessarily have to be single, physical goods, like one glass of orange juice and one glass of lemonade. Usually, they stand for entire bundles, i.e. a combination of several single goods. In this essay, we focus on lotteries, i.e. stochastic consumption. For example,  $X$  may describe a lottery that pays a glass of orange juice with probability  $p = 0.3$  and a glass of lemonade with probability  $1 - p = 0.7$ .

By introducing axioms 3 and 4a (or 4b), it can be shown that rational decision making under risk can be achieved by applying the specific decision rule of (subjective) expected utility maximization. The continuity axiom states that an individual's preference order over two or more lotteries should not be reversed or changed by changes in the probabilities of outcomes, if these changes are sufficiently small. Technically spoken, this means that the preference order does not have any jumps.<sup>7</sup> The independence axiom 4a postulates that the choice between two lotteries should not be influenced by common elements of these two lotteries. Stated differently, the preference relation between two lotteries  $X$  and  $Y$  should not be changed by adding a third lottery  $Z$  to both  $X$  and  $Y$ , as long as the weighting factor  $a$  is the same for both. Then, it must hold true that  $X \succeq Y$  if and only if  $aX + (1 - a)Z \succeq aY + (1 - a)Z$ . Finally, the sure-thing-principle (axiom 4b) applies to ambiguous choices between lotteries where probabilities are unknown and decision makers have to make use of “subjective probability” (Savage (1954), p. 30). It says that, if lottery  $X$  pays strictly more than lottery  $Y$  in at least one state of nature and the same in all remaining states, then  $X$  is preferred to  $Y$ . Put differently, Aumann et al. (2005, p. 2) describe the intuition behind the axiom as follows: “The Sure-Thing Principle ... says that if a decision maker would take a certain action if he knew that an event  $E$  obtained, and also if he knew that its negation  $\sim E$  obtained, then he should take that action even if he knows nothing about  $E$ ”. In particular, it is not necessary to know the probabilities for  $E$  and  $\sim E$ . The decision should be made independent of  $E$  – therefore the sure-thing-principle is sometimes also referred to as an independence axiom for ambiguous choices. We indicate the parallelism of the two by numbering them 4a and 4b.

If all these axioms hold true, this implies that rational decision making can be achieved by maximizing (subjective) expected utility with the preference order being represented by a utility function as introduced by von Neumann and Morgenstern (1944).  $X$  is preferred to  $Y$ , if it yields a higher numerical value of the utility function, or simply a higher utility. If a decision maker is indifferent towards choosing  $X$  or  $Y$  (if  $X \sim Y$ ), both are utility-equivalent. The expected utility of a lottery is simply a linear combination of the utility values assigned to each outcome, weighted with the respective (subjective) probabilities of the outcomes.<sup>8</sup> Note

---

<sup>7</sup> Jumps can be caused by e.g. lexicographic preferences, where single outcomes or properties rather than the whole structure of a certain lottery determine the decision.

<sup>8</sup> To distinguish between risky and ambiguous choices, sometimes the term “lottery” is used only for risky choices with objective probabilities, whereas “gambles” indicate the use of subjective probabilities (see e.g. Binmore (2009)). We follow this terminology in our article.



that a vNM utility function is cardinal, i.e. the preference order deduced from the assigned utility levels remains stable after linear monotone transformations of the utility function. By such a transformation, it is possible to standardize a vNM utility function in a way that the worst outcome in the complete set of potential consequences in a decision problem yields a utility of 0, while the best outcome yields a utility of 1. An advantage of such a utility scale is that utility levels can be interpreted as probabilities: a consequence  $X$  somewhere in between the worst and the best possible outcome yields a utility of, for example, 0.6 when the decision maker is indifferent between  $X$  and a lottery that pays the best possible outcome with a probability of 0.6 and the worst possible outcome otherwise.

A second important advantage, especially in the light of this chapter, is the fact that utilities of different agents within a group can be compared directly by standardizing their utility functions in a way that they yield a utility of 0 in a “worst case event” and a utility of 1 in the “best case event” that everybody in the group can agree on. Assuming that each member of the group behaves in accordance with the axioms introduced above and given that the resulting individual linear expected utility functions can be aggregated in an overall welfare function that is linear itself (see Binmore (1994), p. 281), the group’s decision can be interpreted as resulting from the maximization of expected utility in an analogous way to individual rational agents.

However, there is broad evidence supporting the assumption that the rational homo oeconomicus is not a descriptive but rather a normative model of economic decision making. From the behavioral economics literature, we know that biases and errors can occur. Most notably for the scope of this chapter, there have been several seminal studies on inconsistencies with one or more of the axioms presented above (most prominently the different forms of the independence axiom), therefore questioning the theoretical basis for rational decision making on an individual level. We will present a selection of these problems in section 2.2, namely the Allais Paradox (distinguishing two versions discussed in the literature, the common consequences and the common ratio effect), the Ellsberg Paradox and the Zeckhauser Paradox (also in two different versions).

## 2.2 Paradoxes in Rational Decision Making

### 2.2.1 The Allais Paradox

Maurice Allais (1953) has introduced an experiment that commonly reveals decision making behavior that is inconsistent with the independence axiom and therefore with EUT in general. The test persons are asked to make decisions in two different choice sets, consisting of two lotteries each. There are three outcomes which are common for all four lotteries:  $S1 = 25,000$  MU,  $S2 = 5,000$  MU and  $S3 = 0$  MU.<sup>9</sup> The alternatives only differ in the probabilities that are assigned to these outcomes in each lottery. So far, this is the general setting for “Allais-type” experiments. In the literature, there are two different versions of the Allais paradox that in effect both appeal to the fact that expected utility functions are defined except for monotone linear transformations. The “common consequences effect” makes use of an additive transformation while the “common ratio effect” uses a multiplication by a constant factor.

Let us first consider the common consequences effect as described in the original study by Allais (1953): Lottery  $X$  has the probabilities (0,1,0) for outcomes  $S1$ ,  $S2$  and  $S3$ , respectively, i.e. it pays 5,000 MU with certainty. Lottery  $X'$  has the probabilities (.10, .89, .01). The other pair of lotteries to choose from is  $Y$  with (0, .11, .89) and  $Y'$  with (.10, 0, .90). Frequently, subjects prefer  $X$  over  $X'$  and  $Y'$  over  $Y$ . However, this is not in line with the independence axiom, since  $X \succcurlyeq X'$  implies that the utility of receiving 5,000 MU with certainty, denoted by  $U(5,000)$  is higher than the expected utility of receiving 25,000 MU, 5,000 MU or nothing with probabilities .10, .89 and .01, respectively. Formally, we have

$$U(5,000) > .1U(25,000) + .89U(5,000) + .01U(0)$$

The independence axiom states that this preference structure should persist, regardless of any further elements that are added to both lotteries, likewise – constituting the common consequences. Now, if we add  $.89U(0) - .89U(5,000)$  to both sides of the inequality, we receive

$$.11U(5,000) + .89U(0) > .10U(25,000) + .90U(0)$$

which of course means, that if  $X \succcurlyeq X'$ , it should follow immediately, that  $Y \succcurlyeq Y'$ . However, this is frequently not observable with test persons.

---

<sup>9</sup> In the original study, payoffs are denoted in French Francs. We use a modified setup following Andreoni and Sprenger (2010). In our experiment, we use abstract „monetary units“ (MU) which are converted to EUR at the end.

The common ratio effect works very similar. It is a variation of the original Allais paradox proposed by Kahneman and Tversky (1979)<sup>10</sup>: The probabilities for  $S1$ ,  $S2$ , and  $S3$  are (0,1,0) for Lottery  $X$  and (.98, 0, .02) for Lottery  $X'$ . Lottery  $Y$  has the probability distribution (0, .01, .99) and Lottery  $Y'$  has (.0098, 0, .9902). Note that probabilities for the positive outcomes  $S1$  and  $S2$  are transformed by a multiplication with the factor 0.01, being the “common ratio” of both lotteries. The probabilities of the zero-outcome  $S3$  are the residuals. They need not be transformed, as the utility of  $S3$  is normalized to zero. Again, EUT requires that  $Y \succcurlyeq Y'$  if  $X \succcurlyeq X'$  is observed. However, a reverse choice pattern occurs frequently. Formally, for  $X \succcurlyeq X'$  it must hold that

$$0U(25,000) + 1U(5,000) + 0U(0) > .98U(25,000) + 0U(5,000) + .02U(0)$$

which immediately reduces to  $U(5,000) > .98U(25,000)$  when cancelling the summands that are equal to zero. If, however,  $Y'$  is preferred over  $Y$ , it must hold true that

$$.0098U(25,000) + 0U(5,000) + .9902U(0) > 0U(25,000) + .01U(5,000) + .99U(0)$$

which reduces to  $.0098U(25,000) > .01U(5,000)$ . Multiplying the utility values in this relation with a factor of 100 does not change the preference order of  $Y' \succcurlyeq Y$ , but it is a contradiction to the previous preference order of  $X \succcurlyeq X'$ .

### 2.2.2 The Ellsberg Paradox

Daniel Ellsberg (1961) has formulated a famous choice experiment that provides interesting insights into ambiguity aversion of decision makers. In the lotteries described so far, all outcomes were assigned precise and objective (i.e. common to all decision makers) probabilities. However, sometimes choices have to be made without this information. Since individuals are capable of actually making such decisions, it is assumed that they assign “subjective probabilities” to outcomes in order to operationalize the ambiguous situation. Ellsberg argues that, under the axioms introduced above, revealed preferences over different ambiguous lotteries permit the inference of these subjective probabilities, or at least their relative magnitudes. In other words, if a particular choice is made by an individual, and the axioms above are generally accepted by this individual, then it should be possible to assign certain sets of objective probabilities to the outcomes of the respective lotteries that would induce the individual’s observed decision in a way that conforms to all the axioms. By this

---

<sup>10</sup> Again, we follow the version of Andreoni and Sprenger (2010).

thought construct, the validity of our set of axioms is extended to situations with unknown objective probabilities and SEUT serves as the benchmark for rational decisions. The Ellsberg-paradox, however, describes a situation in which decision makers commonly behave in a way that is irrational, given the deduced subjective probabilities from a previous choice.

Suppose an urn contains 90 balls of three different colors, red, yellow, and black. It is known that 30 balls are red, while the remaining 60 balls are either yellow or black, with an unknown distribution. There are two rounds of playing, with a change in rules after the first choice. In the first round, the decision maker has to decide which gamble she prefers:

- (I) A red ball pays 10,000; yellow and black balls pay 0.
- (II) A black ball pays 10,000; red and yellow balls pay 0.

It is commonly observed that the majority of test persons decide in favor of gamble (I). When constructing subjective probabilities, it is evident that  $(I) \succcurlyeq (II)$  if  $\frac{1}{3} = p^{obj}(red) \geq p^{subj}(black)$ . Alternatively, we can also apply the logic of the “sure-thing principle”: since the event (or state of nature) “yellow” yields the same outcome in both gambles, it should be considered irrelevant for the decision problem. Thus, choosing gamble (I) must stem from a preference relation of the form  $red \succcurlyeq black$ . In the second round, using the same urn with the same (partly unknown) distribution of colored balls, the decision maker has to choose between:

- (III) Red and yellow balls pay 10,000; black balls pay 0.
- (IV) Black and yellow balls pay 10,000; red balls pay 0.

Here, test persons often prefer gamble (IV). Again, the choice should not depend on information about “yellow”, since it yields the same outcome in both gambles anyway. However, with  $p^{obj}(red) \geq p^{subj}(black)$  or simply  $red \succcurlyeq black$ , as deduced from the choice in the first round, it must follow that  $(III) \succcurlyeq (IV)$ . However, it seems that there is an appealing logic behind this seemingly irrational behavior. If a decision maker exhibits a general aversion against ambiguous situations, this leads to a preference order like the one typically observed in the Ellsberg Paradox: the “risky” gamble is always preferred to the “ambiguous” gamble, no matter which subjective probabilities could be deduced from every single choice that is made.

### 2.2.3 The Zeckhauser Paradox

The Zeckhauser Paradox has first appeared in the literature as an example to illustrate the shape of the weighting function in Kahneman and Tversky's (1979) seminal paper introducing the Prospect Theory. The paradox describes a situation where individuals tend to pay more for a reduction of risk than they actually should when deciding rationally in the sense of maximizing expected utility. Suppose an individual is threatened by a pistol pointed at her head which is loaded with one bullet. It has a maximum capacity of six bullets. She can pay an amount  $X$  to remove the bullet from the pistol. The same individual is then again threatened by the same gun, now loaded with four bullets, and can pay an amount  $Y$  to have one of the four bullets removed to increase her chances of surviving. Assuming that being dead – regardless of the amount she has spent to get one bullet removed – constitutes the worst potential outcome  $W$  to the decision maker and being alive after having paid nothing is the best potential consequence  $B$ , we can assign the utility levels of 0 and 1, respectively, to these two states. The question now arises whether the state of having paid  $X$  and being alive in the first setup (denoted by  $C$ ) or having paid  $Y$  and being alive in the second one (denoted by  $D$ ) yields a higher utility – or put differently, whether her willingness to pay is higher for  $X$  or  $Y$  at the margin. With this calibration of the utility function we can interpret the utility of  $C$  as the probability  $p$  that induces indifference between  $C$  or a lottery that yields  $B$  with probability  $p$  and  $W$  otherwise. Thus, it must hold that

$$U(C) = pU(B) + (1 - p)U(W) = p$$

At the margin, she will pay an amount to get the bullet removed from the gun in the first setup that leaves her indifferent between surviving with certainty and having paid  $X$  on the one hand and paying nothing and leaving the bullet in the gun on the other hand. In the latter case, chances to survive are  $\frac{5}{6}$  and the utility in the case of surviving is set to 1. It follows immediately that  $U(C)$  must also be  $\frac{5}{6}$  to induce indifference. The same logic applies to the second setup where  $Y$  is paid to have one of four bullets removed. Here, it must hold that

$$U(D) = pU(B) + (1 - p)U(W) = p$$

Chances to survive are  $\frac{1}{3}$  if no bullet is removed and therefore  $U(D) = \frac{1}{3}$ . At the margin, we see that  $U(C) > U(D)$  and therefore  $C \succ D$ . The consequences  $C$  and  $D$  only differ by the amount of money the individual has left after surviving the shot. It must therefore hold true that  $X < Y$ , regardless of the specific risk attitude (averse, neutral or seeking) that is reflected

in the utility function of the individual. As long as more money provides a higher utility and being alive is preferred over being dead, it is rational to pay strictly more for having one out of four bullets removed than for having the only bullet removed from the gun. For obvious reasons, this setup has not yet been tested experimentally. However, when deciding hypothetically, test persons often report the same numbers for both or even a higher amount for  $X$  than for  $Y$ . There are empirical studies about the monetary valuations of non-financial risks, for example Huber et al. (1987). They find premiums for the complete elimination of health risks that are associated with the usage of chemical products such as insecticides. Aldy and Viscusi (2003) perform a meta-study on mortality and injury risk premia.<sup>11</sup>

We test two versions of the Zeckhauser Paradox. The first one is a thought experiment following the setup described above. The second one is slightly changed in order to induce real incentives with our test persons: We provide test persons with an initial endowment of 10,000 MU. Afterwards, they have to play a game which is similar to Russian Roulette, with the initial endowment at stake (instead of their life). We ask the participants for the maximum amount of money that they are willing to pay for having their chances to keep the 10,000 MU raised by  $\frac{1}{6}$ , resulting in a chance of 100 percent and 50 percent in the two rounds of playing, as above.

### **2.3 Literature Review on Decision Making Paradoxes**

Before we turn to the experimental setup of this study, we provide a brief review on the relevant literature concerning group rationality in the paradoxes that we analyze. In general, social influences on risk preferences are widely discussed in recent years, and a comprehensive review is beyond the scope of this chapter. We recommend the paper by Trautmann and Vieider (2011) for this purpose. Bone et al. (1999) conduct an experimental study of group decision behavior for the Allais paradox in the common ratio form. They find a nearly identical degree of inconsistent choice patterns for individual and group decisions and therefore refute group rationality. However, they neither control for personal attributes of the group members other than their prior individual decision, nor for group-dynamic processes, namely risk shifts that may occur. Furthermore, their groups consist of two persons only. Eliaz et al. (2006) theoretically discuss group decisions and find that the occurrence of the Allais paradox and the phenomenon of choice shifts are caused by the same underlying

---

<sup>11</sup> In case you are wondering, the statistical monetary value of a human life lies somewhere between 4 and 9 million USD.

violation of the independence axiom. Mathauschek et al. (2007) study group rationality in the light of EUT and Portfolio Selection Theory (see Markowitz 1952). Regarding EUT, they analyze common ratio effects, preference reversal effects<sup>12</sup>, and reference point effects.<sup>13</sup> Their study supports group rationality in the sense of Portfolio Selection Theory, but only in part for EUT: the reference point effect tends to occur less frequently in the group treatment, while the other two effects occur equally often. Also in this study, the authors cannot control for preference shifts induced by group formation, as the test persons participate in just one condition, either individually or as a group member. In our experiment, the participants are tested in both conditions. This in turn gives rise to the concern that individual learning effects might confound our results. Bone et al. (1999) do not find evidence for individual learning (in the sense of test persons adopting to the principles of EUT) when repeating the individual treatment after the group discussion. Also with regards to Bayesian updating, which is especially important in the Ellsberg paradox where individuals can learn about the distribution of balls in the non-transparent urn by repeated draws, we are confident that a single repetition of the choice tasks does not provide significant learning opportunities. A study by Trautmann and Zeckhauser (2013) underpins this view. However, to be on the safe side, we randomize the order in which participants are assigned to the individual and the group treatment in our experiment.

As for the Ellsberg Paradox, empirical evidence of group behavior is very limited. Slovic and Tversky (1974) conduct an experiment where small groups of test persons are provided with competing arguments pro and contra violations of the sure-thing principle in the Ellsberg decision task. Afterwards, the participants have to rank the arguments by their persuasive power to them, i.e. they have to evaluate which way of reasoning about the task appeals more logical to them. The vast majority of subjects perceives an argumentation in favor of paradoxical behavior, i.e. contradicting the sure-thing principle, as more convincing and makes their choices accordingly. However, choices are made individually and the authors do not report anything on observations of group polarization. Also, it is not required that the group members agree on a common position.<sup>14</sup> Trautmann et al. (2008) show that ambiguity

---

<sup>12</sup> The preference reversal effect refers to choice patterns, in which subjects sell, say lottery A for a higher price than lottery B, but choose lottery B when it comes to playing one of the lotteries.

<sup>13</sup> The reference point effect states that subjects reverse their valuation of a lottery depending on the presentation as a gain frame or loss frame, i.e. when the algebraic signs of the payoffs are exchanged.

<sup>14</sup> Interestingly, somewhat different results are found in a parallel setting using the Allais paradox in the same experiment. The differences in perceived persuasiveness of the arguments exhibit the same, but weaker tendency towards a reasoning

aversion in Ellsberg-typed choices is strongly driven by social factors, especially the decision maker's fear of being negatively evaluated by peers for preferring an ambiguous over a risky alternative. The paper that is closest to ours is the study by Budescu et al. (2012). They compare individual and group decisions between risky and ambiguous gambles and find higher neutrality towards ambiguity in groups. They also analyze the classical Ellsberg task and find that the percentage of SEUT-consistent choice patterns is not significantly different between groups and individuals. The authors attribute these results to the persuasive power of ambiguity neutrality itself – it is a reasonable compromise. However, they neither control for the perceived persuasiveness of the team members themselves, nor for other social and personal skills. With our experimental design, we take up their conclusion that “obviously, the group interaction does more than simply aggregate the individual opinions” (p. 22). A recent paper by Charness et al. (2013) focusses on the role of persuasive efforts by the group members in ambiguity preference formation. They find that persuasiveness has a significant influence on group decisions by inducing a tendency towards ambiguity neutrality, but only if the test persons are specifically incentivized to persuade their team mates. In our experiment, we do not incentivize any specific social behavior, and indeed, persuasiveness does not drive group rationality in the Ellsberg Paradox, nor does it seem to play a role in preference shifts. However, other skill variables do, and it is fruitful to widen the focus on a broader range of personal and social attributes in order to better understand group decisions under ambiguity. Finally, we could not find any studies on group decision making in the Zeckhauser paradox.

## 2.4 Paradoxes in Group Decisions – Experimental Setup

We extend the literature on differences between individual and team decision making in two aspects. First, we analyze whether personal and social skills influence decision outcomes on both levels, individually and in a group of three. Besides rationality, we also look at potential preference shifts. Second, we are the first to present a comprehensive experimental study of the Zeckhauser Paradox.

The two basic research questions are whether individuals and groups make different decisions under risk and under ambiguity. Concretely, we look at the following:

---

supporting a contradiction of the independence axiom. However, like in our experiment, in actual choices the majority of subjects follows the axiom and shows non-paradoxical behavior in the Allais setting.



**a) Are there any preference shifts caused by switching from individual to group decisions?**

**b) Are teams more rational than individuals?**

These questions have already partly been answered in the literature, as discussed in the previous section. Therefore, in a next step, we widen the scope of the analysis by addressing two new research questions:

**c) Is rationality influenced by personal and social skills?**

**d) Are preference shifts influenced by personal and social skills?**

To answer our research questions, we conduct an experimental study with two treatments and a final questionnaire.<sup>15</sup> In the “individual” treatment, the test persons independently make their decisions between several sets of lotteries that are designed to detect paradoxical behavior on the individual level, namely the Allais Paradox, the Ellsberg Paradox, and the Zeckhauser Paradox. Furthermore, we estimate the individual participants’ preferences towards risk by applying a variation of the Becker/DeGroot/Marschak (BDM) method (Becker et al. 1964). For 24 risky lotteries, the test persons are asked to report selling prices for which they are indifferent between playing the lottery and selling it. Second, in order to estimate the individual degree of ambiguity aversion, we apply the method of Halevy (2007) by converting the 24 risky lotteries into ambiguous gambles, i.e. using the same payoff structure with unknown probabilities. Again, the test persons report the selling prices for the gambles at which they are indifferent between playing and selling them. The difference between the two selling prices for the risky and the ambiguous version of each lottery reflects the degree of ambiguity aversion. Participants are instructed that it is rational to report their “true” selling price.<sup>16</sup> The payoffs in the BDM lotteries and the paradoxes range from 0 to

---

<sup>15</sup> Appendix 2.A, Appendix 2.B, Appendix 2.C, and Appendix 2.D contain the instructions and the tasks from both treatments. Appendix 2.E contains the final questionnaire. We recommend reading these appendices right after completing this section before continuing with the next section.

<sup>16</sup> See Appendix 2.B for the instructions that subjects received. We must admit that the selling prices (i.e. the certainty equivalents for the lotteries) that we elicited are potentially biased due to a small but theoretically crucial error that we detected months after the experiments were run. The original BDM procedure has the advantage that it is designed in a way that makes it incentive compatible for subjects to report their „true“ selling price, i.e. not to under- or overstate it. In the original version, subjects report their selling price and receive an offer for the lottery afterwards. If the offer is higher than their demanded selling price, they are paid the offer price instead of playing the lottery. Thus, from the perspective of the seller of the lottery ticket, the BDM procedure has the same properties as a two-person sealed-bid *second-price* auction (also called „Vickrey Auction“, see Vickrey (1961)): In order to sell the lottery, one’s own selling price has to be lower than the offer price, which can be imagined as a concurring bid of another player. In such an auction, it is always optimal to report one’s true valuation. The intuition behind this is that the price that will be paid for the lottery is independent of one’s personal valuation: the lowest bidder wins the auction, but pays a price equal to the second-lowest bid (i.e. the offer price). In the instructions for our experiment, we mistakenly stated that subjects can sell their lottery ticket for the *selling price* in case that

10,000 MU, thus we can infer that the test persons' attitudes towards risk and ambiguity deducted from the lotteries and gambles actually reflect their preference structure in the relevant domain. The only exception is the highest per capita payoff in the Allais Paradox (25,000 MU). We wanted this number to be perceived as very high in order to grasp the intention of the original Allais task and did so by choosing a payoff that is clearly above the range of payoffs that the subjects get used to evaluate during the course of the experiment.

In the "group" treatment, participants are put together in teams of three persons and run through the same questionnaire structure like in the individual treatment to control for any changes in tendencies to paradoxical choices as well as preferences towards risk and ambiguity that may be caused by the formation of the group. Note that the actual choice sets differ in both treatments. First, payoffs in the group treatment are multiplied by the number of team members to keep expected payoffs per capita constant. Second, the lotteries used in the BDM procedure are computed dynamically in both treatments, i.e. the payoffs in later lotteries depend on the valuations made in earlier stages of each treatment. The experiment is conducted as a within subject design, i.e. all test persons take part in both treatments. In order to overcome potential complications caused by interaction effects between both treatments (e.g. learning or sensitization to perceived dependencies between both trials (Greenwald 1976)), we also apply a crossover design: half of the participants take the individual treatment first; the other half starts with the group treatment. The questionnaire is always answered at

---

it is lower than the offer price. This makes the elicitation procedure formally equivalent to a two-person sealed-bid *first-price* auction. In such an auction, the optimal bidding strategy is *not* to report one's true valuation, because the expected payoff from the auction can be increased by stating a higher selling price. Maskin and Riley (2000) provide an overview of optimal bidding strategies in first-price auctions for a broad set of assumptions. Subjects have to trade off the higher payoff conditional on selling the lottery against the lower probability that the transaction is made because it gets more unlikely that the offer price is above the selling price the higher the latter is set. Thus, if subjects behave perfectly in accordance with auction theory, selling prices should generally be higher in our elicitation procedure than if they were elicited with the original BDM procedure. However, there is also theoretical and empirical evidence that the consequences of our mistake are limited for the purpose of our experiment. First, if biases occur, it is reasonable to assume that they affect individual and group assessments similarly. Thus, they do not influence the analysis of risk shifts. Second, the Revenue Equivalence Theorem of auction theory (see Riley and Samuelson (1981)) says that even if bidding behavior differs, the expected revenue to the seller is identical in first-price and second-price auctions if offer price and selling price are independent from each other. This is the case in our experiment, because the offer price is determined by a random mechanism that is independent from the risk preference of the subject and vice versa. Third, following the argumentation of Maskin and Riley (1984), the amount by which agents overstate their true valuation in the selling price depends on their risk preference in a way that risk averse subjects overstate less than risk neutral subjects. In our special case where the auctioned good is a lottery, risk averse agents by definition report lower selling prices than risk neutral persons even when stating their valuations truthfully. Thus, applying a first-price auction to elicit preferences might even yield a higher discriminatory power than the original BDM procedure. This constitutes an interesting field for further research on risk elicitation mechanisms. Finally, Rutström (1998) finds that despite their theoretical equivalence, subjects bid significantly higher in Vickrey Auctions than in BDM Auctions when buying a privately-valued good. Translating this to our experiment, where subjects bid to *sell* the good (the lottery), we would expect that the selling prices elicited by the BDM procedure are lower than in a Vickrey Auction. While it is unclear which of these two mechanisms actually elicits selling prices that are closer to the true valuations, we take her finding as clear evidence that actual behavior in an experiment deviates strongly from theoretical predictions. Therefore, even though the confusion of selling price and offer price in the instructions has strong theoretical implications on behavior, we are confident that the subjects reported their true certainty equivalents as their required selling prices, as they were extensively instructed to do so.

the end of the experiment. To ensure incentive compatibility, we randomly choose three lotteries and determine the payoff for each participant. Azrieli et al. (2012) show in a theoretical analysis on experimental methods that choosing one lottery for payment instead of playing out all lotteries is the only incentive compatible design, at least under a little restrictive set of assumptions about the subjects' preferences. The intuition is that subjects otherwise will integrate choices over multiple tasks and this is likely to distort behavior in single choices. For example, when choosing a lottery with outcomes 0 and 10,000 MU, subjects should expect to leave the experiment with one of the two outcomes. However, when they get paid for many of these lotteries, the final amount instead converges to the expected value of the lottery. We believe that, given the large number of choices in our experiment, randomly choosing three lotteries for payment is still incentive compatible. Furthermore, smoothing outcomes a little bit diminishes the risk that frustrated subjects who did not win anything will not participate when recruited for other experiments. Thus, the subject pool is sustained. The final MU amounts are converted to Euros by multiplication with the factor 0.00071429. This factor has been chosen ex post in order to pay a sum of 15.00 EUR for the average total gain.<sup>17</sup>

In the final questionnaire, we survey gender, age, nationality, familiarity with the tested paradoxes, acceptance of the axioms, and personal and social skills. Each test person is asked to evaluate a set of attributes of her team members along scales for analytical competence, goal-orientation, persuasiveness, and leadership shown in the group treatment. As alternative measures for personal and social skills, we also have three consolidated expert judgments made in the admission process of the Bachelor program at hand. We also use the final grades from the respective secondary schools as an additional cognitive skill measure.

The peer evaluations among the team members are made on a seven points scale ranging from -3 to +3 with the lowest value indicating a very weak perception of the respective attribute and the highest value a very strong characteristic. For each participant and each attribute, we calculate the average score that has been awarded by the two peers in order to achieve the variables "Analytical competence", "Goal-orientation", "Persuasiveness", and "Leadership".

The admission process for the Bachelor of Science program in Management at WHU – Otto Beisheim School of Management takes place three months prior to the begin of the first

---

<sup>17</sup> This amount corresponds to the usual salary that undergraduate students could have earned in two hours when being employed with the university.

semester of studies on the university campus. When the experiment took place, all participants were enrolled in this program. During the process, each participant has been evaluated by seven independent and trained referees and in four different settings, namely a 50-minutes group discussion with five other applicants on a topic given shortly before the start (observed and evaluated by three referees), a presentation with subsequent moderation of another discussion on the prepared topic of the presentation (two referees), two single in-depth interviews (one referee per interview), and a multiple choice test on analytic skills in which the applicants are asked to interpret diagrams (not refereed; answers were compared mechanically with the sample solution). We do not use the scores achieved in the single interviews, as these were mainly about the applicants' motivation for the program and the field of study as well as curricular and extracurricular activities which we do not consider relevant for our analysis. Each referee assigns a score from 1 to 10 (with 10 being the best) to each applicant in each part of the admission process. The referees are instructed to make use of the full range of the scale in order to clearly differentiate the qualities of the applicants. The results of the multiple choice tests are also converted into scores from 1 to 10. For each participant in the experiment, we calculate the average (over the respective numbers of referees) scores for the group discussion and the presentation. Thus, we end up with three additional variables that describe the skills of the test persons, namely "Group discussion", "Presentation", and "Diagram task". In both, the group discussion and the presentation, the referees are instructed to evaluate the communication skills, as well as the clearness, stringency and creativity of the produced arguments. However, in the group discussion, there is also an additional strong focus on team-orientation and leadership skills that distinguishes the two measures. The diagram task is a pure measure of quantitative analytical skill.

## **2.5 Sample**

Our sample consists of 108 individuals who finished the experiment. 54 of them started with the individual treatment and were randomly assigned to 18 groups afterwards, while the other half of the participants started with the group treatment. All participants were undergraduate students in business administration. 27.8 percent of the participants (30 subjects) were female. The age of the test persons ranged from 18 to 24 years. 105 of the participants were German, with 4 of them having a second nationality. The remaining 3 persons came from other European countries. All participants were fluent in German.

We asked all the participants in the final questionnaire whether they are familiar with the three paradoxes that are used in the experiment. The choice tasks in the experiment were not named after or referred to the original paradoxes, so if a participant answers that she is not familiar with one of them, we can infer that she did not recognize it from the experimental tasks and that her choices are not influenced by her knowledge about choice patterns that are consistent with EUT or SEUT. While the Zeckhauser Paradox is usually not a part of introductory lectures or textbooks in Microeconomics, Decision Theory, or Finance, the Allais Paradox and the Ellsberg Paradox are much more popular and often serve as illustrative examples for irrational (i.e. not (S)EUT-consistent) behavior. This is reflected by the numbers: 19.4 percent of the subjects (21 persons) answered that they are familiar with the Allais Paradox. 15.7 percent (17 persons) said that they know the Ellsberg Paradox, but only 2.8 percent (3 persons) have heard of the Zeckhauser Paradox.

Furthermore, we asked each subject whether she agreed with the five basic axioms that are fundamental to EUT and SEUT, namely the completeness axiom, the transitivity axiom, the continuity axiom, the independence axiom, and the sure-thing principle. We presented a verbal description of each axiom's implication for choice behavior and asked the subjects if they agreed with this. If this is true for all axioms for any subject, EUT and SEUT choice predictions should serve as normative benchmarks for this subject. If all axioms but the sure-thing principle are accepted, EUT can be regarded as a normative choice model for risky environments, but SEUT does not hold for ambiguous choices. Table 2.1 summarizes the given answers.

Table 2.1: Acceptance of EUT and SEUT Axioms

	<b>completeness</b>	<b>transitivity</b>	<b>continuity</b>	<b>independence</b>	<b>sure-thing</b>
<b>I agree</b>	85 (78.7%)	98 (90.7%)	73 (67.6%)	75 (69.4%)	62 (57.4%)
<b>I disagree</b>	7 (6.5%)	4 (3.7%)	14 (13.0%)	15 (13.9%)	28 (25.9%)
<b>Not sure</b>	16 (14.8%)	6 (5.6%)	21 (19.4%)	18 (16.7%)	18 (16.7%)

Note: percentages of total sample reported in parentheses

The completeness and the transitivity axiom are the most fundamental ones. They propose that subjects can make binary choices from a set of alternatives and that the resulting overall preference order is consistent. They are widely accepted by the subjects in our experiment. The continuity axiom implies that preferences can be mapped into continuous utility

functions. More than two thirds of our subjects accept this axiom. The theoretical role of the independence axiom is to allow for linear combinations of utility values, which is important when calculating expected utility values for lotteries. The sure-thing principle modifies the independence axiom to cases of ambiguity by comparing states of nature without any assigned probabilities. We observe that the acceptance of the sure-thing principle is clearly lower than of the independence axiom. Furthermore, the sure-thing principle has the highest number of subjects that explicitly disagree. We take this as a signal to be careful with interpreting SEUT predictions as normative.

To conclude the overview on our sample data, we provide some descriptive statistics on the personal and social skill measures that we use. The peer evaluations in analytical competence, goal-orientation, persuasiveness, and leadership, are surveyed in the final questionnaire. The data on the participants' final grade in secondary school, as well as the results from the assessment center during the selection progress for the study program, were provided by the program administration.<sup>18</sup> We see that the subjects in the sample are rated on almost the full scale that is available for each measure, with the mean values always being slightly skewed towards the “better” side.

Table 2.2: Peer Evaluations, Final Grade from Secondary School, and Assessment Center Scores

	Mean	Std. Deviation	Minimum	Maximum
<b>Analytical competence</b>	1.486	1.049	-3.0	3.0
<b>Goal-orientation</b>	1.440	0.988	-2.0	3.0
<b>Persuasiveness</b>	1.005	1.065	-2.5	2.5
<b>Leadership</b>	0.718	1.049	-2.5	3.0
<b>Final grade (School)</b>	1.499	0.372	1.0	2.7
<b>Diagram task (AC)</b>	5.940	1.818	2.0	10.0
<b>Group Discussion (AC)</b>	5.993	1.571	1.3	8.7
<b>Presentation (AC)</b>	6.463	1.715	2.0	10.0

---

<sup>18</sup> With the enrollment in the study program, the students have agreed that their data may be used anonymously for research purposes.

## 2.6 Results

### 2.6.1 Elicited preferences towards risk and ambiguity

We start our analysis by answering research question a), i.e. we want to find out if and how preferences differ between individual and group decisions. To measure preferences towards risk and ambiguity we follow the methods of Becker et al. (1964) and Halevy (2007). Our test subjects (individuals and groups) assign selling prices to risky lotteries and ambiguous gambles. From these observations and the properties of the lotteries we can infer preferences. To elicit risk preferences we use lotteries with known payoffs and known probabilities. If test person  $i$  (or group  $j$ ) makes a discount on the expected value of the lottery  $k$  ( $EV_k$ ) when stating her selling price ( $SP_{i,k}^{risk}$ ), this reflects risk aversion, while a premium over the expected value shows risk seeking. For the ambiguous gambles, we use the same payoffs as in the risky lotteries, but the probabilities are unknown. Thus, the selling prices for the ambiguous gambles ( $SP_{i,k}^{amb}$ ) have to be compared to those reported for their risky counterparts to elicit ambiguity preferences. We calculate the discounts that determine risk preferences ( $r$ ) and ambiguity preferences ( $a$ ) of individual  $i$  (and group  $j$  accordingly) in lottery  $k$  as follows.

$$r_{i,k} = \frac{SP_{i,k}^{risk} - EV_k}{EV_k}$$

$$a_{i,k} = \frac{SP_{i,k}^{amb} - SP_{i,k}^{risk}}{SP_{i,k}^{risk}}$$

To have a first indication of the observable preferences at hand, we determine the overall risk preference ( $r^*$ ) for individual  $i$  (and group  $j$ ) by calculating the mean over the 24 risky lotteries in our experiment  $r_i^* = \sum_{k=1}^{24} r_{i,k}$ . The overall ambiguity preference is estimated accordingly. We will refine these estimations later on. On average, individuals as well as groups are slightly risk averse, but neutral towards ambiguity. The sample means (standard deviations in parentheses) are shown below.

Table 2.3: Risk and Ambiguity Preferences for Individuals and Groups

	Individuals	Groups
<b>Risk Preference <math>r^*</math></b>	-0.053 (0.111)	-0.041 (0.073)
<b>Ambiguity Preference <math>a^*</math></b>	0.012 (0.181)	-0.009 (0.176)

Note: The resulting preference parameters  $r_i^*$  and  $a_i^*$  are averaged across all individuals; the resulting preference parameters  $r_j^*$  and  $a_j^*$  are averaged across all groups. The standard deviations of the respective sample means are reported in parentheses.

A Kolmogorov-Smirnov test reveals that the observations for both variables  $r^*$  and  $a^*$  are sufficiently close to being normally distributed. Two dependent t-tests do not reveal any statistical significance in the differences between the average discounts of individuals and groups for neither risk preferences nor ambiguity preferences. We can also analyze the relationship between an individual's (a group's) risk and ambiguity preferences by calculating the Pearson correlation coefficients of the preference parameters in both samples. Neither of the coefficients is significantly different from zero. Thus, the risk preference parameters do not correlate with the ambiguity preference parameters.

We categorize an individual (a group) as risk averse if  $r_i^* < -0.055$ . For  $-0.055 < r_i^* < 0.055$  we attest neutrality towards risk, and for values of  $r_i^* > 0.055$ , it is classified as risk seeking. The same thresholds are applied for the categorization of ambiguity preferences. The two tables below report risk and ambiguity preferences that we observe in our sample. For convenience, we report in parentheses the number of individuals that are represented by the respective group preferences.

Table 2.4: Categorized Risk Preferences for Individuals and Groups

Risk Preference	Individuals	Groups
<b>averse</b>	50	14 (42)
<b>neutral**</b>	40	19 (57)
<b>seeking**</b>	18	3 (9)

Note: In the "Groups" column, the numbers in parentheses indicate the number of individuals that are represented by the groups in each risk preference category. \*\*\*, \*\*, and \* indicate a statistically significant change in the observed frequency of the respective preference between individual and group decisions on the 99%, 95%, and 90% confidence level, as computed in a McNemar test for differences in dependent and binary variables.



Table 2.5: Categorized Ambiguity Preferences for Individuals and Groups

Ambiguity Preference	Individuals	Groups
<b>averse</b>	33	10 (30)
<b>neutral</b>	22	10 (30)
<b>seeking</b>	53	16 (48)

Note: In the “Groups” column, the numbers in parentheses indicate the number of individuals that are represented by the groups in each ambiguity preference category. \*\*\*, \*\*, and \* indicate a statistically significant change in the observed frequency of the respective preference between individual and group decisions on the 99%, 95%, and 90% confidence level, as computed in a McNemar test for differences in dependent and binary variables.

A  $\chi^2$ -Test does not indicate a significant relationship between the categorized variables (“averse”, “neutral”, “seeking”) for risk and ambiguity preferences in the individual sample. The same result is obtained for group decisions. However, the  $\chi^2$ -Test is likely to yield imprecise results due to the small sample size for groups. As seen before, from observing either an individual’s or a group’s risk preference, we cannot conclude to their ambiguity preference. Trautmann and van de Kuilen (2014) provide an overview of many studies that analyze the interplay of risk and ambiguity preferences. They stress that those studies who – like ours – do either find no or even negative correlations, usually use similar measures like this study, namely the differences between the certainty equivalents of risky lotteries and correspondent ambiguous gambles. The standard finding in the literature is according to their review a positive correlation, but then risk and ambiguity preferences are elicited independently from each other.

For both, risk and ambiguity preferences, we observe an overall tendency to neutrality when comparing group decisions to the corresponding individual decisions. Only 40 individuals are categorized as risk neutral, while the groups that behave risk neutrally represent 57 subjects. This is a significant change, as indicated by a McNemar test for differences in binary, dependent observations. Migrations from risk averse and risk seeking subjects contribute equally to this higher number of risk neutral group members. However, because of the higher relative change in risk seeking preferences, only this reduction is statistically significant. For ambiguity preferences, we observe the same pattern, but less pronounced. Indeed, none of the changes are statistically significant when applying the McNemar test. Ambiguity seeking is the modal preference for both, individuals and groups.

However, to further analyze the effect of group formation on preferences, it is not meaningful to compare the observed risk and ambiguity preference parameters  $r_i^*$  and  $a_i^*$  for individuals

with their direct counterparts for groups, as the choices in the BDM procedure are path-dependent and therefore in general not identical for the individual and the group treatments. Instead, we estimate simple one-parameter linear utility functions for the relevant domain of monetary (per capita) values in the BDM lotteries that can easily be compared across individuals and groups, even though they are not derived from exactly the same choice sets.

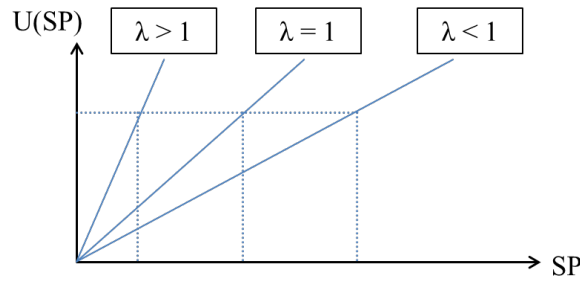
For risky choices (with known probabilities), under the assumptions of EUT, it is possible to assign utility values to each lottery before the subjects provide their selling prices. This is true because utility functions are cardinal, i.e. they can be linearly and monotonically transformed without implying changes in preferences. For example, multiplying the utility values of all available choices with the same factor or adding a constant to all of them does not change the rank order. Therefore, it is convenient to standardize a utility function in a way that it yields a utility value of 0 for the worst option that is available and a value of 1 (or any multiple of 1) for the best option. By assuming that a payment of 0 MU is the worst outcome for every subject and a payment of 10,000 MU is the best outcome, we can easily assign utility values of 0 and 10,000 to these outcomes, respectively. Therefore, the expected utility of a lottery with the outcomes 0 MU and 10,000 MU must be the probability-weighted average of these two utility values. For example, the first lottery in our experiment has an expected utility of 5,000. In the next step, the selling price that an individual or a group assigns to this lottery elicits their risk preference. If for any given decision maker the utility of the expected value of a lottery is equal to the expected utility of this lottery, this is commonly known as risk neutrality. That means that the decision maker is indifferent between playing the lottery (that has an expected utility of 5,000), and receiving the expected monetary value of this lottery (5,000 MU) as a sure payment. If a sure payment that is lower than the expected monetary value of the lottery yields the expected utility of the lottery, this indicates risk aversion. For example, if the decision maker states a selling price of 3,000 MU for the lottery, she applies a discount to the expected monetary value and would still be indifferent between receiving 3,000 MU and playing the lottery that yields an expected utility of 5,000. Therefore, for this person, a sure payment of 3,000 MU must also yield a utility of 5,000. For risk seekers, the sure payment must be higher than the expected monetary value of the lottery to induce indifference.

In total, we have 24 BDM lotteries with pre-specified utility values that are evaluated by each individual and each group.<sup>19</sup> For each individual, we estimate the risk preference parameter  $\lambda_i$  in the following simple utility function by performing an OLS regression with the constant set to 0.

$$U(SP_i^{risk}) = \lambda_i SP_i^{risk}$$

If  $\lambda_i = 1$ , individual  $i$  is risk neutral. For  $\lambda_i > 1$ , the individual is risk averse, and for  $\lambda_i < 1$ , it is risk seeking.<sup>20</sup>

Figure 2.1: Risk Preference Parameters of Linear Utility Functions



Analogously, we estimate the risk preference parameters  $\lambda_j$  for the group decisions. Furthermore, by pooling the 3x24 individual decisions of each group member, we estimate a “hypothetical” group risk preference parameter  $\lambda_j^{hyp}$ . This is the risk preference parameter that would be deducted from the group decisions if there were no other effects from group formation than averaging individual preferences. Thus, we define the difference between the actual and the hypothetical risk preference parameter as the preference shift that occurs due to group decision making.

$$\Delta\lambda_j = \lambda_j - \lambda_j^{hyp}$$

<sup>19</sup> See Appendix 2.C for an overview.

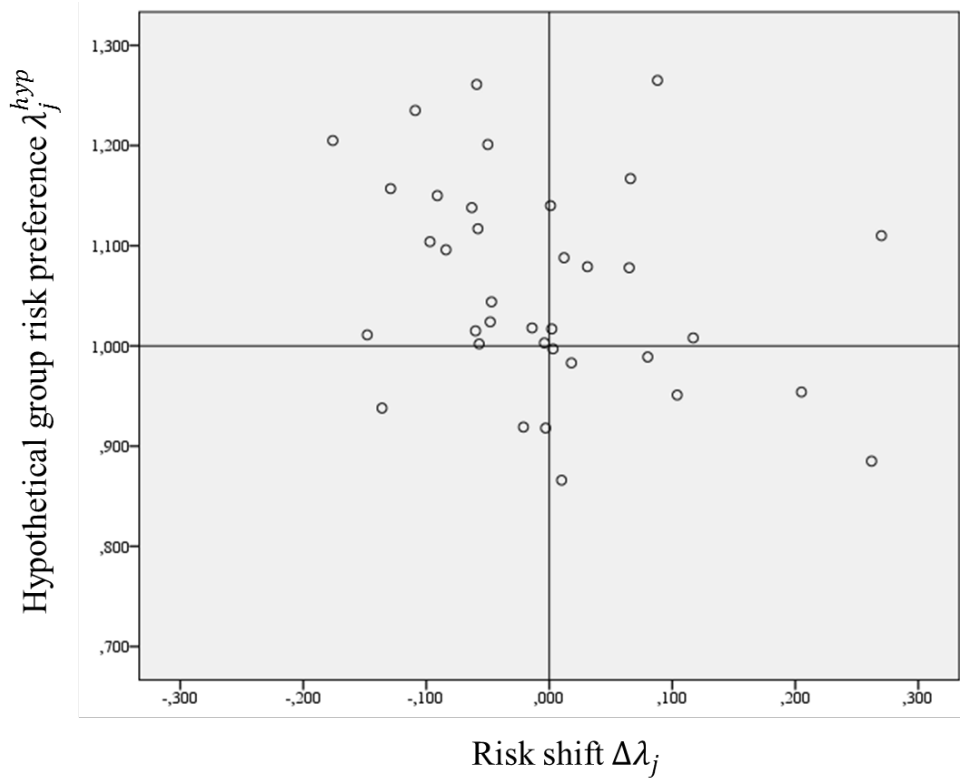
<sup>20</sup> As a quick robustness check, we calculate the means of  $\lambda_i$  for the subsamples the subjects that are categorized as risk averse, risk neutral, and risk seeking by the critical values of  $r^*$ . We find that risk averse subjects have a  $\lambda_i$  of 1.238 (standard deviation 0.154), risk neutral subjects have a  $\lambda_i$  of 1.014 (std. dev. 0.038), and risk seeking subjects have a  $\lambda_i$  of 0.875 (std. dev. 0.092).

For  $\Delta\lambda_j > 0$ , the observed group decisions exhibit higher risk aversion (or less risk-taking) than implied by the individual decisions.

The group risk preference parameter  $\lambda_j$  ranges from 0.802 to 1.380 and has a mean of 1.056 and a standard deviation of 0.124. The hypothetical group risk preference parameter  $\lambda_j^{hyp}$  has a range between 0.866 and 1.265, with a mean of 1.059 and a standard deviation of 0.105. Kolmogorov-Smirnov tests do not indicate significant deviations from normal distributions for both variables ( $\lambda_j$ :  $Z(36) = 0.567$ ,  $p = 0.905$  and  $\lambda_j^{hyp}$ :  $Z(36) = 0.789$ ,  $p = 0.563$ ).

The average risk shift across all 36 groups is -0.003 and has a standard deviation of 0.105. Also in this case, the Kolmogorov-Smirnov test does not indicate a significant deviation from a normal distribution ( $Z(36) = 0.848$ ,  $p = 0.469$ ). Therefore, we can confirm by a t-test that the average risk shift is not significantly different from zero ( $t(35) = -0.191$ ,  $p = 0.849$ ). The risk preferences that are elicited in group decisions are on average the same – showing a moderate degree of risk aversion – as those reflected in the individual decisions. However, with values ranging from -0.176 to 0.27, we do in fact observe considerable preference shifts within most of the groups, and the comparison of the minimum and maximum values, as well as standard deviations, suggests that the actual groups' risk preferences are more dispersed than the hypothetical group estimates, with the mean being preserved. The Pearson correlation coefficient for  $\lambda_j^{hyp}$  and  $\Delta\lambda_j$  is -0.304 and is significant on the 90% confidence level ( $p = 0.071$ ). This suggests that those groups who consist (on average) of highly risk averse subjects, become slightly more risk seeking and vice versa. The scatterplot shown below also shows that those groups that have a negative shift parameter  $\Delta\lambda_j$ , indicating higher risk seeking or less risk aversion, tend to start from a hypothetical group risk preference  $\lambda_j^{hyp}$  above one, therefore consisting of primarily risk averse group members. Of those (overall fewer) groups that are on average risk seeking ( $\lambda_j^{hyp} < 1$ ) in the first place, the majority exhibits a positive risk shift parameter, which means that they become less risk seeking or even risk averse.

Figure 2.2: Hypothetical Group Risk Preference and Risk Shift



For ambiguity preferences, the analysis is not so straightforward. By definition, the probabilities of the lottery outcomes are unknown. Thus, we cannot calculate the utility values ex ante and therefore cannot estimate utility functions for ambiguous choices. Instead, we take up the idea of SEUT and use implicit subjective probabilities to overcome this problem. The subjective probability assigned to payoff  $X$  in a given gamble is the probability value that would justify a subject's or a group's selling price for the gamble if it were a risky lottery instead, given their estimated utility function under risk. Remember that subjects evaluate lotteries (with known probabilities) and gambles (with unknown probabilities) pairwise in a way that for each lottery, after stating their selling price, they are asked to state their selling price for another lottery "with the same payoffs as before, but this time with unknown probabilities". Imagine a risk neutral subject ( $\lambda_i = 1$ ) that reports the expected value of 5,000 as her selling price for a lottery that pays  $X = 10,000$  MU with a 50 percent chance and  $Y = 0$  MU otherwise facing the corresponding gamble. Suppose she lowers her selling price to 2,500 MU. Being risk neutral, it is straightforward to see that this would be her valuation if the lottery above had a probability of 25 percent for paying 10,000 MU. Thus, the subjective probability that can be deducted from this selling price for the gamble is 25 percent.

Now imagine the same lottery being evaluated by a risk averse subject ( $\lambda_i = 1,5$ ). For her, the selling price of the lottery would be 3,333 MU. However, she also quotes a selling price of 2,500 MU for the appendant gamble. Do the two subjects have the same attitude towards ambiguity?

To calculate the implicit subjective probability that is underlying the selling price for an ambiguous gamble for a subject with a given risk attitude, we need to solve the following equation, with  $U(SP^{amb}) = \lambda_i SP^{amb}$ ,  $U(X) = \lambda_i X$ , and  $U(Y) = \lambda_i Y$ .

$$U(SP^{amb}) = p^{subj}U(X) + (1 - p^{subj})U(Y)$$

The equation states that the utility of the selling price for the ambiguous gamble must equal the expected utility of a corresponding lottery that features the probabilities  $p^{subj}$  and  $1 - p^{subj}$ . Thus, we restate the ambiguous decision problem as a risky choice, which comes with the crucial advantage that we can use the previously estimated utility functions to describe individuals' and groups' attitude towards ambiguity.<sup>21</sup>

Solving the equation for  $p^{subj}$  for the risk neutral subject from the example above yields indeed a subjective probability of 25 percent. The risk averse subject that stated the same nominal selling price, however, has implicitly applied a subjective probability of only 16.6 percent. Comparing these subjective probabilities to the objective probability of the original lottery, we see that the risk averse subject in this example has a higher degree of ambiguity aversion than the risk neutral subject, as she has applied a higher discount on the objective probability.<sup>22</sup> We define the conversion factor that subject  $i$  applies to the objective probability for payoff  $X$  in lottery  $k$  when stating her selling price for the corresponding ambiguous gamble as follows.

$$\gamma_{i,k} = \frac{p_{i,k}^{subj}}{p_{i,k}^{obj}}$$

---

<sup>21</sup> To elicit ambiguity preferences, there are basically two approaches used in the literature. Subjects either report evaluations for ambiguous gambles or they adjust the objective probabilities of a lottery until they are indifferent between an ambiguous gamble and the lottery. This second method is often referred to as matching probabilities or probability equivalents. Our approach links the reported selling prices of individuals and groups to their implicit subjective probabilities.

<sup>22</sup> Also if both subjects had applied a proportional discount of 50 percent on their individual selling prices for the risky lottery, the subjective probabilities that were implied by this are not identical. In the example, the risk averse subject would cut down her valuation by half when applying a subjective probability of 11 percent.

For group decisions, we again use an index  $j$  analogously. In total, each individual and each group quotes 22 selling prices for ambiguous gambles.<sup>23</sup> We report the average discount factors  $\gamma_i$  and  $\gamma_j$  by leaving out the lottery index  $k$ .<sup>24</sup> For convenience, we also report the implicit subjective probabilities that are congruent with individual and group selling prices under their respective risk preference parameters. Standard errors are reported in parentheses.

Table 2.6: Probability Conversion Factors and Subjective Probabilities implied by Individual and Actual Group Decisions

	$\gamma_i$	$\gamma_j$	$p_i^{subj}$	$p_j^{subj}$
<i>all gambles</i>	1.003 (0.018)	1.000 (0.033)	-	-
$p^{obj} = 0.25$	1.782 (0.028)	1.753 (0.051)	0.445 (0.007)	0.438 (0.013)
$p^{obj} = 0.5$	0.830 (0.026)	0.847 (0.055)	0.415 (0.013)	0.424 (0.027)
$p^{obj} = 0.75$	0.674 (0.015)	0.652 (0.020)	0.505 (0.017)	0.489 (0.022)

The average probability conversion factors are very close to 1.000 for both, individuals and groups. This suggests neutrality towards ambiguity and confirms our first tentative estimations of  $\alpha^*$ . However, we observe that when splitting the sample by the underlying objective probabilities of the lotteries, conversion factors differ substantially from 1.000. For both, groups and individuals, the implied subjective probabilities are around 44 percent when the objective probability is 25 percent, while in the cases where the objective probability is the highest with 75 percent, subjective probabilities are adjusted downwards to around 50 percent. This cannot be explained by any generally applicable attitude towards ambiguity, but it rather suggests a simple tendency to the middle, which is especially pronounced in the latter case with the highest objective probabilities. However, when the objective probability is actually 50 percent, neither groups nor individuals report selling prices for the ambiguous gambles that would imply 50 percent as an underlying subjective probability. Instead, the average subjective probabilities for these gambles are even somewhat lower than for those with a 25 percent objective probability. These results suggest that preferences towards

<sup>23</sup> Lotteries 1 and 2 did not have ambiguous counterparts, as the nominal payoffs for the first three lotteries are identical. Thus, all gambles would have looked similar.

<sup>24</sup> Note that the resulting term for  $p^{subj}$  is mathematically not defined if  $X = Y$  and subjects (rationally) report  $X$  as their selling price. We observe this in six individual choices and in one group choice. In these cases, we manually set the subjective probabilities equal to the objective probabilities.

ambiguity depend on the objective probabilities of lotteries that serve as reference points or priors. The conversion function is seemingly non-monotonic and not symmetric around the middle point of 50 percent, but with only three discrete probability values that we use in this experiment, we refrain from drawing any further conclusions about the specific functional form. Trautmann and van de Kuilen (2014) conclude in their meta-study that this is indeed a typical pattern that can also be found in other experiments, too. It would be interesting to derive the conversion function for a quasi-continuous range of objective probabilities in order to understand better how decisions under ambiguity are made. We leave this question to further research.

Because the probability conversion factors and the respective implied subjective probabilities reflect the individual and group-specific risk preferences, we can compare these parameters between individuals and groups. In analogy to our analysis of risk shifting, we first calculate a hypothetical probability conversion factor  $\gamma_j^{hyp}$  for each group in our sample by simply averaging the three individual  $\gamma_i$  from each group member. The “ambiguity shift” is then calculated as the difference between the actual and the hypothetical conversion factors and subjective probabilities, respectively.

$$\Delta\gamma_j = \gamma_j - \gamma_j^{hyp}$$

$$\Delta p_j^{subj} = p_j^{subj} - p_j^{subj\ hyp}$$

Table 2.6 already shows that probability conversion factors for individuals and groups are very close to each other. Therefore, the hypothetical parameter estimates are quite similar to the actual values, and neither the group shift in the conversion factor  $\Delta\gamma_j$  nor the shift in subjective probabilities  $\Delta p_j^{subj}$  is significantly different from zero. This holds for aggregate values as well as for subsamples split by objective probabilities.<sup>25</sup> Thus, on average, we do not observe any preference shifts between groups and individuals when making decisions under ambiguity, similar to what we observed in the case with risk preferences. When looking at the correlations between hypothetical subjective probabilities or probability conversion rates and the respective shift parameters, we find that there is no relationship for the overall sample. We only find a highly significant Pearson correlation coefficient of -0.462 (p=0.001) in the subsample of those gambles for which the associated objective probability is 25

---

<sup>25</sup> Tables are available from the authors upon request.



percent. For the other subsamples, there are no systematic relationships between these two variables.<sup>26</sup> Thus, for ambiguity preferences, we cannot in general attest a convergence of preferences as we did for risk preferences where groups tended to adjust towards neutrality. This is in line with the results of Keller et al. (2007) who test for differences in ambiguity preferences between individuals and teams of two. Keck et al. (2012) on the other hand find a slight tendency of teams of three towards neutrality. Charness et al. (2013) state that ambiguity neutrality has “a persuasive edge over ambiguity seeking [...] and to a lesser degree over ambiguity averse subjects” (p.13). In section 2.6.6 we discuss risk and ambiguity preference shifts in more detail by relating them to social and personal skills.

## 2.6.2 Individual and Group Rationality in the Allais Paradox

Research question b) is answered separately for each of the three paradoxes. We start by looking at individual and group rationality in the Allais Paradox. The results for both versions of the Allais Paradox are striking in the sense that they hardly reflect any paradoxical behavior. As for the common consequences version, only three of the individuals and none of the groups showed inconsistent behavior. All others have decided rationally in the sense of EUT. Furthermore, all participants have chosen the second option in each pair of lotteries, i.e.  $X'$  and  $Y'$ . Accordingly, in all groups the majority of individual members have decided consistently and all group decisions reflect the preference of the majority.

Table 2.7: Individual Majorities and Group Decisions on Group Level in the Allais Paradox (common consequences,  $N = 36$ )

AP_CC	Consistent (Majority)	Inconsistent (Majority)	Total
Consistent (Group)	36	0	36
Inconsistent (Group)	0	0	0
Total	36	0	36

The picture for the common ratio version looks similar. Only one individual decision is inconsistent, all others have chosen the lotteries  $X'$  and  $Y'$ , with the choices made by the respective groups being identical. The group with the inconsistent individual reported that

---

<sup>26</sup> It is sufficient to look at the subjective probabilities directly, because they are linear transformations of the probability conversion rates and therefore the correlation coefficients are identical.

they could not agree on a common preference and therefore accepted the default payoff of 0 MU for this choice.

Table 2.8: Individual Majorities and Group Decisions on Group Level in the Allais Paradox (common ratio, N = 36)

AP_CR	Consistent (Majority)	Inconsistent (Majority)	Total
Consistent (Group)	35	0	35
Inconsistent (Group)	1	0	1
Total	36	0	36

Thus, the common explanation for paradoxical behavior in the Allais settings – namely the certainty effect – does not apply for our sample. Individual and group choices both reflect a very clear tendency to maximize the expected value. This is not to be confounded with pure risk-neutrality in the sense that there is no preference for certainty at all. However, moderate risk-aversion as measured for our sample with the BDM lotteries obviously does not justify giving up considerable gains in expected value for the sake of certainty. We see three potential explanations for this uncommon result. First, there might be a selection bias in the test persons. All our test persons were enrolled in a Bachelor program in Business Administration and it can be argued that their educational background as well as certain demographics, their personality and interests are systematically different from the population average. Second, the experimental setup itself might have fostered rational behavior by providing fairly extensive learning opportunities in the sense that the test persons are likely to develop a more thought-out risk attitude when evaluating a large number of lotteries compared to a short “one shot” experiment that captures more spontaneous decisions. Third, the original version of the Allais Paradox deals with very large payoffs (0, 1 million and 5 million French Francs compared to 0, 5,000 and 25,000 MU which translated into 0, 3.57 and 17.86 Euros) and therefore decisions might be framed fundamentally different. When using very large payoffs in an experimental study, it is prohibitively expensive to install real incentives, i.e. to pay out the actual payoffs. Furthermore, it seems much more natural that the certainty effect sets in when the certain amount is 1 million rather than 3.57 Euros.

### 2.6.3 Individual and Group Rationality in the Ellsberg Paradox

We continue by answering research question b) for the Ellsberg Paradox. We also present the results for research question c) for individual decisions in this section. We omitted this part in the previous section on the Allais Paradox due to the lack of variation in the obtained data. The effect of personal and social skills on group rationality (i.e. the discussion of research question c) on the group level) will be analyzed later in section 2.6.5 for both, the Ellsberg and the Zeckhauser Paradox. In the Ellsberg Paradox, there are in principle two ways to determine whether choice behavior is consistent. First, we analyze whether choices conform to the predictions by SEU Theory. This is the classical approach that Daniel Ellsberg (1961) chooses to make the case for preferences against ambiguity. We then analyze in a second step, whether the elicited ambiguity preferences from the BDM lotteries conform to the choices made in the Ellsberg task.

Only 26 of 108 individuals (7 out of 36 groups) make consistent choices in the light of SEU Theory, i.e. they either choose lotteries *I* and *III* (13 persons, 1 group) or lotteries *II* and *IV* (13 persons, 6 groups). 70 individuals (26 groups) demonstrate choice behavior that is in line with ambiguity aversion by preferring lotteries *I* and *IV*. 12 test persons (3 groups) show a preference for ambiguity by choosing lotteries *II* and *III*. When looking at the preferences of the majority and the actual group decision<sup>27</sup>, the picture looks as follows. While in 26 groups an inconsistent choice pattern is supported by the majority of the individuals, and in 4 groups a majority of consistent choices also translates into a consistent group decision, there are also 3 groups with a majority of inconsistent individual choices but a consistent group decision, and 3 more groups with a majority of consistent individual choices but an inconsistent group decision. Thus, there seems to be no positive effect of group formation on rationality in the sense of SEU Theory, as the small number of three cases where an inconsistent majority of individuals actually achieves a rational outcome is exactly offset by three other cases where the group effect goes in the opposite direction. When translating the group decisions into the number of individuals behind them, there is even a slight tendency of higher irrationality in groups: while 26 individual decisions are consistent with SEU Theory, the 7 groups which decide rationally represent only 21 test persons.

---

<sup>27</sup> Davis and Hinsz (1982) point out that in binary group choices the preference of the majority of the group members is crucial for choice behavior and potential choice shifts.

Table 2.9: Consistency with SEUT for Majorities of Individuals and Group Decisions on Group Level in the Ellsberg Paradox (N = 36)

EP - SEUT	Consistent (Majority)	Inconsistent (Majority)	Total
Consistent (Group)	4	3	7
Inconsistent (Group)	3	26	29
Total	7	29	36

Looking at the majorities and the resulting group decisions yields a fairly good indication whether group formation increases or lessens consistency in choices, but it is imprecise as we do not take into account the exact numbers of individuals who decide rationally or irrationally in a group. Take for example a group that is classified as having a consistent majority. This could be two or three individuals, and in turn, there could be one or no inconsistent individual in this group. In order to achieve statistical tractability of the effect of group formation on rationality, we assign two dummy variables to each individual that indicate consistency (=1) or inconsistency (=0) in the individual and the group decision. For all members of a specific group, the values of the group dummy are of course identical. We then run a McNemar test for differences in dependent and binary variables.<sup>28</sup> The biggest part of all individuals (71 persons) decides inconsistently on their own and in the group, while 11 test persons make consistent choices in both treatments. The number of individuals that decide rationally on their own, but irrationally in the group is slightly higher than individuals whose inconsistent choice behavior becomes consistent when being in a group (16 versus 10), but this difference is statistically insignificant.

Table 2.10: Consistency with SEUT of Individual and Common Group Decisions on Individual Level in the Ellsberg Paradox (N = 108)

EP - SEUT	Consistent (Group)	Inconsistent (Group)	Total
Consistent (Individual)	11	16	27
Inconsistent (Individual)	10	71	81
Total	21	87	108

Note: A McNemar test for differences in dependent binary observations does not indicate a statistically significant difference in consistency between individual and group decisions (N = 108,  $\chi^2 = 0.962$ ,  $p = 0.327$ ).

---

<sup>28</sup> We also apply this procedure in the two versions of the Allais Paradox. There is no statistically significant difference in rationality between individuals and groups in any of the two versions. We do not report these results separately, as the figures on group choices and majorities already show this very clearly. Tables are available from the authors upon request.

In line with the previous literature on choice under ambiguity, we can infer from the results above that SEU Theory is a bad predictor for choice behavior in the Ellsberg Paradox. Choices are predominantly inconsistent, with no significant difference between groups and individuals.

Instead, groups and individuals care about ambiguity, i.e. the reasoning of SEU Theory where decision makers use identical subjective probabilities in both Ellsberg tasks does not seem adequate to describe choice behavior. Still, it is questionable whether we see higher degrees of rationality in both, individual and group decisions, when we compare the choices in the Ellsberg task with the preferences towards ambiguity that we elicit in the BDM gambles.

Subjects and teams are classified as ambiguity averse, neutral, or seeking depending on their parameter values of  $\gamma_i$  and  $\gamma_j$ , respectively. We define a value range between 0.95 and 1.05 as representing neutrality towards ambiguity. Lower values indicate ambiguity aversion; higher values occur in the case of ambiguity seeking. Depending on the specific attitude towards ambiguity, different choice patterns in the Ellsberg task are rational. When applying the elicited ambiguity preferences calculated over all gambles as rationality criterion, we see that both, individuals and groups make more consistent choices than in the SEUT case above. Now, 14 groups choose rationally, with only half of them consisting of a majority of rational individuals – but there are also 5 teams that consist of predominantly rational individuals who still cannot agree on a consistent group choice pattern. Overall, there are still 22 teams who make inconsistent choices, and, as seen in the SEUT case, the largest cluster in the table below is located in the lower right, indicating that most groups consist of two or even three irrational subjects who are unable to come to a rational group decision. Even though the quota of teams with consistent group preferences who have only one or no rational individual but still agree on a rational group decision is slightly higher than in the SEUT case, we cannot conclude that groups make clearly more rational decisions than individuals when applying general ambiguity preferences as a rationality criterion instead.

Table 2.11: Consistency with Elicited Ambiguity Preferences (measured by  $\gamma_i$  and  $\gamma_j$  over all gambles) for Majorities of Individuals and Group Decisions on Group Level in the Ellsberg Paradox (N = 36)

EP-BDM, all gambles	Consistent (Majority)	Inconsistent (Majority)	Total
Consistent (Group)	7	7	14
Inconsistent (Group)	5	17	22
Total	12	24	36

In fact, when looking at the individual decisions and comparing them one by one with the corresponding group decision, it becomes obvious that individual decisions are even significantly more rational than group decisions: there are 32 individuals that make consistent choices on their own but inconsistent choices in their teams, while there are only 12 test persons that decide inconsistently as individuals but rational as group members. This difference is highly significant, according to a McNemar test for differences binary and dependent variables.

Table 2.12: Consistency with Elicited Ambiguity Preferences (measured by  $\gamma_i$  and  $\gamma_j$  over all gambles) of Individual and Common Group Decisions on Individual Level in the Ellsberg Paradox (N = 108)

EP – BDM, all gambles	Consistent (Group)	Inconsistent (Group)	Total
Consistent (Individual)	9	32	41
Inconsistent (Individual)	12	55	67
Total	21	87	108

Note: A McNemar test for differences in dependent binary observations indicates a statistically significantly higher degree of consistency in individual decisions on the 99% confidence level (N = 108,  $\chi^2 = 8.205$ ,  $p = 0.004$ ).

We have shown before that both, individual and group ambiguity preferences differ substantially across the associated objective probabilities. Therefore, we also categorize the groups' and individuals' ambiguity preferences using gambles with identical objective probabilities. Using the gambles with associated fifty-fifty chances only yields the best descriptive model of choice behavior in the Ellsberg task with 18 consistent groups and 21

groups who consist of two or more rational individuals.<sup>29</sup> Group choices seem to be rational to a slightly lesser degree than individual choices, since there are only three groups that decide rationally even though the minority of team members is rational on an individual basis, while there are six groups that decide irrationally, even though at least two out of three members are individually rational.

Table 2.13: Consistency with Elicited Ambiguity Preferences (measured by  $\gamma_i$  and  $\gamma_j$  over gambles with  $p^{obj} = 0.5$ ) for Majorities of Individuals and Group Decisions on Group Level in the Ellsberg Paradox (N = 36)

EP-BDM, $p^{obj} = 0.5$	Consistent (Majority)	Inconsistent (Majority)	Total
Consistent (Group)	15	3	18
Inconsistent (Group)	6	12	18
Total	21	15	36

However, applying the McNemar test to the individual observations demonstrates that there is no significant difference in rationality between group decisions and individual decisions.

Table 2.14: Consistency with Elicited Ambiguity Preferences (measured by  $\gamma_i$  and  $\gamma_j$  over gambles with  $p^{obj} = 0.5$ ) of Individual and Common Group Decisions on Individual Level in the Ellsberg Paradox (N = 108)

EP-BDM, $p^{obj} = 0.5$	Consistent (Group)	Inconsistent (Group)	Total
Consistent (Individual)	35	21	56
Inconsistent (Individual)	19	33	52
Total	54	54	108

Note: A McNemar test for differences in dependent binary observations does not indicate a statistically significant difference in consistency between individual and group decisions (N = 108,  $\chi^2 = 0.025$ ,  $p = 0.874$ ).

Overall, we conclude that in the Ellsberg Paradox, group decisions are clearly not more rational than individual decisions. Depending on the criterion that we apply to define consistency of choice behavior with preferences, groups even do significantly worse. We see

---

<sup>29</sup> We do not report results for the other objective probabilities. The results for the 0.75 sample are almost identical to the fifty-fifty case. When using objective probabilities of 0.25 only, practically all teams and individuals are ambiguity seeking, but most of them make ambiguity averse choices in the Ellsberg task. This is clearly the model with the lowest descriptive power.

that the overall highest degree of consistency in individual and group decisions in the Ellsberg Paradox is attained when ambiguity preferences are measured exclusively over gambles with an associated objective probability of 0.5, with no significant difference in individual and group rationality. When we use all gambles to determine ambiguity preferences, this yields significantly less consistent group decision patterns compared to individual decisions.

In order to understand how ambiguity preferences on the one hand and personal and social skills on the other hand influence choice behavior, we run a number of logistic regressions on the individual level. The dependent variables are dummies which indicate rational choice behavior in the Ellsberg task according to SEUT, to the preference parameters  $\gamma_i$  calculated over all gambles, and  $\gamma_i$  for all gambles with  $p^{obj} = 0.5$ , respectively. As independent variables, we use all peer evaluations from the final questionnaire, the scores from the admission test, a dummy variable that indicates whether a subject has taken the individual treatment first, a dummy variable that indicates whether the subject is familiar with the Ellsberg Paradox, a gender dummy, five dummy variables that indicate acceptance of the axioms fundamental to EUT and SEUT as described in the final questionnaire, and also two dummies that indicate ambiguity aversion or ambiguity seeking as measured in the binary BDM gambles. These preference dummies are determined to match the applied rationality criterion in each regression, i.e. we use gambles with an associated objective fifty-fifty chance only if the applied criterion is the ambiguity preference in this domain. In the case where consistency with SEUT is analyzed, we also use  $\gamma_i$  for gambles with  $p^{obj} = 0.5$  only to determine the value of these dummies. Neutrality towards ambiguity always serves as the base category.

The results for rational choice in terms of SEUT and according to the elicited ambiguity preferences are all very clear. The entire regression model does not have any significant explanatory power when SEUT is the rationality criterion.<sup>30</sup> Interestingly, the dummy that indicates familiarity with the Ellsberg Paradox produces odds ratios significantly higher than one in all specifications, which suggests that the subjects are more likely to make an SEUT-consistent choice in this case. Given that the paradoxical nature of the Ellsberg task is precisely due to the fact that SEUT is violated, this specific knowledge might cause the subjects to decide in accordance with SEUT. The fact that the dummies for ambiguity

---

<sup>30</sup> We test several specifications with different subsets of the independent variables. None of them yields sufficient explanatory power at the usual confidence levels, as indicated by the  $\chi^2$  statistics. Tables are available from the authors upon request.



aversion and ambiguity seeking do not produce odds ratios that statistically differ from one (they should both be below one if ambiguity neutrality was the real reason for SEUT conforming choices) supports this notion. However, we are very careful with this conjecture, as the regression model as a whole does not support this.

When using the individual ambiguity preferences as measured over all gambles and exclusively over those with  $p^{obj} = 0.5$ , none of the control variables have any significant influence on choice behavior, except for the ambiguity aversion dummies. This holds true for all specifications of the regression models.<sup>31</sup> The odds ratios for ambiguity aversion are significantly higher than one – indicating that this specific individual preference determines choice consistently. Besides statistical significance, the level of the produced odds ratios shows that there is a very clear distinction between consistent and non-consistent decision makers in the Ellsberg task and this is moderated by ambiguity aversion: when using  $\gamma_i$  over all gambles, subjects who are classified as ambiguity averse are about five times more likely than ambiguity neutral or seeking subjects to make a consistent choice. And when using the  $\gamma_i$  that are derived from gambles with an objective fifty-fifty chance, the odds ratio even increases to nine. This result supports the original notion by Daniel Ellsberg (1961), namely that ambiguity aversion explains choice behavior. Interestingly, we do not see ambiguity neutral and ambiguity seeking subjects making individually consistent choices to such a high extent, i.e. both odds ratios are not statistically significantly different from one.

#### 2.6.4 Individual and Group Rationality in the Zeckhauser Paradox

We now present the results for research question b) (group rationality versus individual rationality) and research question c) (potential influences of personal and social skills on rationality) on the individual level for the two versions of the Zeckhauser Paradox. The group level will be discussed in section 2.6.5.

##### 2.6.4.1 Russian Roulette Version

In the classical Zeckhauser Paradox, the willingness to pay under EUT for having the only bullet removed from the gun should be strictly lower than for having one of four bullets removed. 18 test persons and 9 groups report their WTP in line with EUT. In 3 groups, the majority of individuals report consistent WTPs that also translate into a consistent group

---

<sup>31</sup> Tables are available from the authors upon request. All specifications have significant explanatory power at the 10 percent confidence level or higher, as indicated by the  $\chi^2$  statistics.

decision. The remaining 6 groups with consistent answers have a majority of individual members that are inconsistent. 25 groups are inconsistent by the majority of the members and also report inconsistent group WTPs. Furthermore, there are 2 groups that are constituted with a consistent majority but give inconsistent group answers. Thus, we see a slight tendency for groups to make more rational choices in the Russian Roulette Version of the Zeckhauser Paradox, but inconsistent WTPs prevail across most groups and individuals.

Table 2.15: Individual Majorities and Group Decisions on Group Level in the Zeckhauser Paradox (Russian Roulette, N = 36)

<b>ZP_RR</b>	<b>Consistent (Majority)</b>	<b>Inconsistent (Majority)</b>	<b>Total</b>
<b>Consistent (Group)</b>	3	6	9
<b>Inconsistent (Group)</b>	2	25	27
<b>Total</b>	5	31	36

This difference is insignificant on the usually required minimum confidence level of 90 percent, as the McNemar test indicates. However, the resulting p-value is only slightly above 0.1.

Table 2.16: Consistency of Individual and Common Group Decisions on Individual Level in the Zeckhauser Paradox (Russian Roulette, N = 108)

<b>ZP_RR</b>	<b>Consistent (Group)</b>	<b>Inconsistent (Group)</b>	<b>Total</b>
<b>Consistent (Individual)</b>	9	9	18
<b>Inconsistent (Individual)</b>	18	72	90
<b>Total</b>	27	81	108

Note: A McNemar test for differences in dependent binary observations does not indicate a statistically significant difference in consistency between individual and group decisions (N = 108,  $\chi^2 = 2.370$ , p = 0.124).

The task in the Zeckhauser Paradox is different from the other two paradoxes in a way that not only a binary choice is reported (i.e. paying more or less than before), but the test persons are asked for a monetary valuation of a reduction in a given risk. This allows us to look closer at the group dynamics. Let  $WTP_{i,j}^{4bullets}$  and  $WTP_{i,j}^{1bullet}$  be the willingness to pay of individual  $i$  ( $i = 1, 2, 3$ ) in group  $j$  ( $j = 1, \dots, 36$ ) for the removal of one bullet in the “4 bullets” and the “1 bullet” task of the classical Zeckhauser Paradox, respectively. We

calculate the sum of the individual WTPs for each group in both tasks, take the difference, and obtain the total amount that each group would pay more in the “4 bullets” scenario if they would simply determine their group decision by summing up their individual WTPs.<sup>32</sup> We compare this hypothetical amount with the difference of the actual group decisions  $WTP_j^{4bullets}$  and  $WTP_j^{1bullet}$  and take the average over all groups:

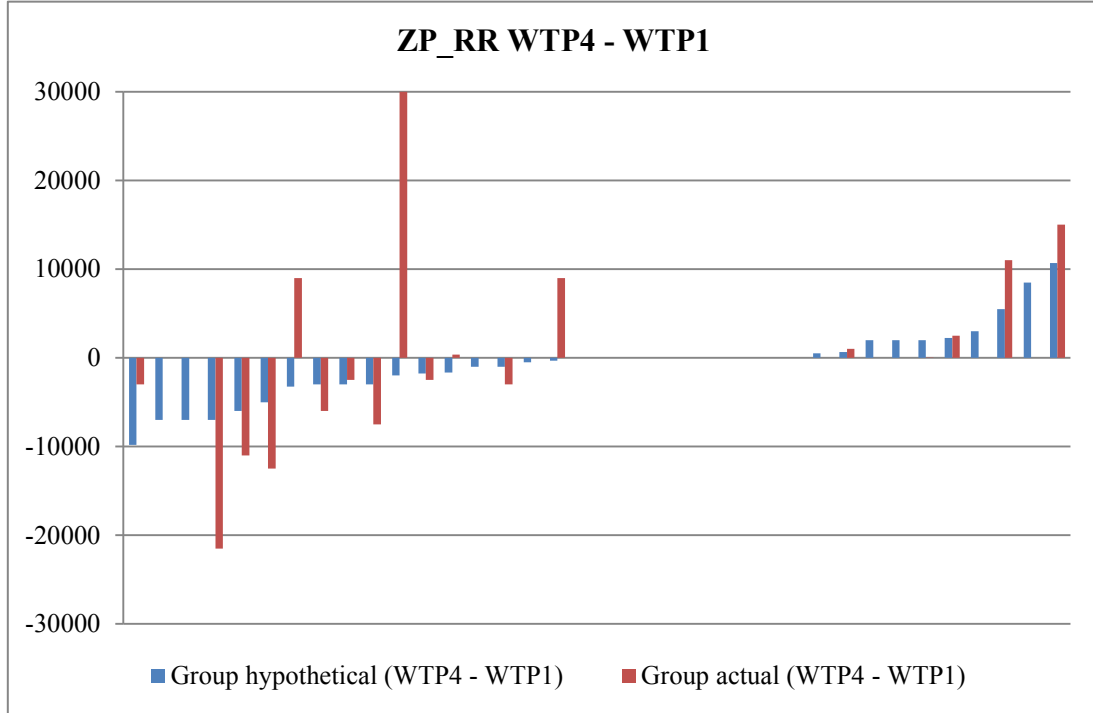
$$\frac{1}{36} \sum_{j=1}^{36} \left[ (WTP_j^{4bullets} - WTP_j^{1bullet}) - \sum_{i=1}^3 (WTP_{i,j}^{4bullets} - WTP_{i,j}^{1bullet}) \right] = 934$$

When considering the hypothetical average group choice  $\frac{1}{36} \sum_{j=1}^{36} \sum_{i=1}^3 (WTP_{i,j}^{4bullets} - WTP_{i,j}^{1bullet})$ , groups would pay 701 MU less in the 4bullets task and thereby behave irrationally. However, in the actual average group choice  $\frac{1}{36} \sum_{j=1}^{36} (WTP_j^{4bullets} - WTP_j^{1bullet})$ , the WTP is 233 MU higher in the 4bullets task and therefore in line with predictions of EUT. On average, we observe a group shift of 934 MU in the relative valuations that leads to seemingly more consistent behavior in the group treatment than in the individual treatment. However, this result is largely driven by very few exceptional cases. The following graph depicts the hypothetical and actual differences in the willingness to pay in the 4 bullets and the 1 bullet case for each team. The median difference for both, hypothetical and actual differences in valuations is zero. There are only four cases where the hypothetical difference is negative (indicating inconsistent valuations), but the actual difference is positive. Interestingly, there are 18 teams that agree on identical payments for the removal of a bullet in both scenarios, while only 9 teams would end up doing this if they would behave according to their hypothetical valuations. Thus, group decisions do not generally increase consistency with EUT, but nevertheless, there is a clear tendency to avoid differences in valuations across the two situations.

---

<sup>32</sup> Recall that the total amount of MU in the game is 30,000 in the group treatment and 10,000 in the individual treatment in order to keep the stakes per capita constant to rule out potential changes in behavior stemming from different individual absolute risk exposure. To compare both decisions, we therefore have to consider the sum of individual WTPs rather than the average.

Figure 2.3: Hypothetical vs. Actual Differences in the Willingness to Pay (Zeckhauser Paradox, Russian Roulette version)



We also look at the implicit absolute changes in risk tolerance in the Zeckhauser Paradox that might arise from group formation. A difference between the actual and the hypothetical WTP for risk reduction (or complete avoidance in the one bullet case) informs us about potential risk shifts in the loss domain.<sup>33</sup> Even though we do observe substantial differences between hypothetical and actual WTPs on average for both, the one bullet and the four bullets case, a Wilcoxon signed rank test does not indicate a significant difference (one bullet case:  $Z = -1.245$ ,  $p = 0.213$ ; four bullets case:  $Z = -0.216$ ,  $p = 0.829$ ). The average differences in WTP per group are as follows:

$$\frac{1}{36} \sum_{j=1}^{36} \left[ (WTP_j^{1bullet}) - \sum_{i=1}^3 (WTP_{i,j}^{1bullet}) \right] = -1738$$

and

$$\frac{1}{36} \sum_{j=1}^{36} \left[ (WTP_j^{4bullets}) - \sum_{i=1}^3 (WTP_{i,j}^{4bullets}) \right] = -803$$

<sup>33</sup> Recall that we did not observe any significant risk shifts for groups in the BDM lotteries. However, we know from Prospect Theory that preferences can differ in the loss domain and the gain domain. This might also be true for groups.

As described for the Ellsberg Paradox in the previous section, we also run a binary regression to analyze the potential influence of personal and social skills, risk preferences (as measured in the BDM lotteries), recognition of the Zeckhauser Paradox, acceptance of the fundamental axioms, gender, and the order in which a subject completed the treatments, on rational behavior in the Russian Roulette version. We test various specifications with different sets of independent variables, but none of them has significant explanatory power.<sup>34</sup>

#### 2.6.4.2 Dice Game Version

The results for our modified version of the Zeckhauser Paradox look as follows: 51 individuals and 20 groups report values that are consistent with EUT. In 13 cases, a consistent group majority translates directly into a consistent group valuation and in 12 cases, an inconsistent group decision comes with an inconsistent majority of individuals. However, there were also 7 group decisions that were consistent even with the majority having made inconsistent valuations and also 4 groups that made inconsistent choices with actually consistent individuals by the majority. Thus, in the Dice Game, the fraction of rational WTP constellations is generally higher than in the Russian Roulette Version, and again, there is a small advantage for groups, with seven groups being able to decide rationally with only a minority of rational team members, compared to four teams where a rational majority cannot agree on a rational group outcome.

Table 2.17: Individual Majorities and Group Decisions on Group Level in the Zeckhauser Paradox (Dice Game, N = 36)

<b>ZP_DG</b>	<b>Consistent (Majority)</b>	<b>Inconsistent (Majority)</b>	<b>Total</b>
<b>Consistent (Group)</b>	13	7	20
<b>Inconsistent (Group)</b>	4	12	16
<b>Total</b>	17	19	36

However, once again this difference is not statistically significant, as a McNemar Test reveals.

---

<sup>34</sup> Tables are available from the authors upon request.

Table 2.18: Consistency of Individual and Common Group Decisions on Individual Level in the Zeckhauser Paradox (Dice Game, N = 108)

ZP_DG	Consistent (Group)	Inconsistent (Group)	Total
Consistent (Individual)	37	14	51
Inconsistent (Individual)	23	34	57
Total	60	48	108

Note: A McNemar test for differences in dependent binary observations does not indicate a statistically significant difference in consistency between individual and group decisions (N = 108,  $\chi^2 = 1.730$ , p = 0.188).

When looking at the actual values for WTP, we see the following. The notation follows the same logic as above with *1pip* and *4pips* indicating the maximum number of pips on the die that lead to the loss of the initial endowment. In the average hypothetical group choice  $\frac{1}{36} \sum_{j=1}^{36} \sum_{i=1}^3 (WTP_{i,j}^{4pips} - WTP_{i,j}^{1pip})$ , groups would pay 1581 MU more in the 4pips task, which is in line with EUT. Thus, contrary to the Russian Roulette version, individual valuations are on average consistent, even though only slightly less than half of the participants (51 of 108) report consistent individual WTPs. When comparing actual and hypothetical differences between the WTPs in both cases, the actual difference for groups is on average a little bit *smaller* (-172 MU) than the average of the team members, but WTP constellations for groups are still consistent with EUT.

In absolute terms, as seen before in the Russian Roulette version of the task, groups have a lower WTP for risk reduction in both cases, indicating a lesser degree of risk aversion in the loss domain:

$$\frac{1}{36} \sum_{j=1}^{36} \left[ (WTP_j^{1pip}) - \sum_{i=1}^3 (WTP_{i,j}^{1pip}) \right] = -1471$$

and

$$\frac{1}{36} \sum_{j=1}^{36} \left[ (WTP_j^{4pips}) - \sum_{i=1}^3 (WTP_{i,j}^{4pips}) \right] = -1643$$

In the Dice Game version, the Wilcoxon test finds significant differences with  $Z = -2.441$ , p = 0.015 in the one pip case and  $Z = -3.006$ , p = 0.003 in the four pips case.

Once again, we also analyze the influence of our comprehensive set of control variables on rational behavior of individuals in the Dice Game version. This time, the male dummy

produces odds ratios significantly lower than one in all model specifications, indicating that men are significantly less likely to show EUT consistent WTPs in this version of the Zeckhauser Paradox than women.<sup>35</sup> For both versions of the Zeckhauser Paradox, we find that risk preferences do not explain rational behavior, in opposition to the previously discussed Ellsberg Paradox, where ambiguity aversion fully explains choice behavior.

## 2.6.5 Group Rationality and Personal and Social Skills

In accordance to the multivariate analyses on the individual level, we now want to see if and how rational decisions, even though they are not more common on the group level, are driven by risk and ambiguity preferences as well as personal and social skills. While the measurement of the group preferences is straightforward – we use the group preference parameters  $\lambda_j$  and  $\gamma_j$  and the group shift parameters  $\Delta\lambda_j$  and  $\Delta\gamma_j$  – the question arises how we can measure the personal and social skills of a whole team. In principle, there are three possibilities. First, we could assume that a team is only as good as its weakest member in a specific skill. Second, the best available skill value among the team members could be at the same time the one for the whole group. And third, it could be a mix of all individual scores, for example the average. We do not know which concept is the most accurate one to determine the “true” skill value of a team. It could also be the case that for different skills in question, different concepts apply, or that the best way to measure a team’s skill level depends on the nature of the task that has to be solved. Therefore, we will follow all three approaches and will discuss their feasibility after we see the results.

Remember that for the Ellsberg Paradox, we have used three different methods to determine whether a decision is rational or consistent: SEUT, ambiguity preferences measured over all gambles, and ambiguity preferences measured over gambles with an objective probability of fifty percent only. Furthermore, we will now apply all three approaches to measure the group skill variables (average, best individual score, worst individual score) to these three rationality models so that we end up with nine different regressions for the Ellsberg Paradox. In the Zeckhauser Paradox, we only have one method to classify decisions as rational, but since we have tested two different versions (Russian Roulette and Dice Game), we have six different

---

<sup>35</sup> Again, we do not report the detailed regression results. Tables are available from the authors upon request.

regressions here. For the sake of brevity, we do not report all the results in detail, but we will focus on those models and variables that have significant explanatory power.<sup>36</sup>

For the Ellsberg Paradox, one model specification yields significant explanatory power to determine group rationality under SEUT, and two models can explain group rationality according to the measured ambiguity preferences over all gambles and over gambles with an objective probability of fifty percent, respectively, as measured by the Wald  $\chi^2$ -statistics for logistic regressions.

To explain group rationality under SEUT, the usage of the average scores of the group members yields overall significance of the logistic regression model, while the scores of the best or the worst subject do not. Remember that for individual decisions our regression model failed to explain rationality under SEUT. When applying the ambiguity preferences measured over all gambles to determine consistency, the models that use the scores of the best and the worst team member in a specific skill have significant explanatory power, while the one that uses the average skill values does not. And finally, when using the ambiguity preferences derived from the gambles with an objective probability of fifty percent only, using the mean values as well as the scores of the best individual in each team yields an altogether significant explanatory power, but using the worst individual scores does not. Again, remember that, even though our regression models for individual consistency of ambiguity preferences in the Ellsberg Paradox had significant explanatory power, the only significant variables were the preference parameters themselves. In group decisions, consistency with SEUT and ambiguity preferences is also significantly influenced by other factors, especially personal and social skills. The following table shows the results of the five significant regression models.

---

<sup>36</sup> All other regression tables are available from the authors upon request.



Table 2.19: Determinants of Consistency of Group Decisions in the Ellsberg Paradox (N = 36)

	SEUT	$\gamma_j$ over all gambles		$\gamma_j$ over gambles with $p^{obj} = 0.5$	
	Average individual	Best individual	Worst individual	Best individual	Average individual
<b>Treatment Dummy (1 = individual treatment first)</b>	0.0126* (0.076)	0.0130** (0.024)	0.7212 (0.800)	0.3990 (0.754)	0.4180 (0.541)
$\gamma_j$	3.33 (0.835)	79.06 (0.609)	246.97 (0.311)	$5.80 e^{-15***}$ (0.006)	$2.78 e^{-5*}$ (0.056)
$\Delta\gamma_j$	354143.20* (0.080)	250491.90 (0.160)	106.63 (0.476)	$2.70 e^{-19**}$ (0.043)	0.0056 (0.456)
<b>Analytical Competence</b>	0.0075* (0.061)	0.0163*** (0.007)	1.73 (0.324)	0.0123** (0.034)	0.1030 (0.417)
<b>Goal Orientation</b>	3.87 (0.579)	68.47** (0.019)	0.4869 (0.426)	$319106.9***$ (0.005)	12.53 (0.101)
<b>Persuasiveness</b>	0.0334 (0.199)	0.0287** (0.033)	2.88 (0.132)	1.97 (0.694)	1.2008 (0.892)
<b>Leadership</b>	29.30 (0.160)	0.3037 (0.576)	$0.2302**$ (0.032)	$1.84 e^{-4***}$ (0.005)	0.5455 (0.680)
<b>Final Grade Secondary School</b>	1.45 (0.917)	$1559.63^*$ (0.063)	0.1867 (0.425)	$1.18 e^{-6**}$ (0.039)	0.0564 (0.361)
<b>Group Discussion</b>	2.68 (0.432)	1.33 (0.843)	0.8194 (0.776)	$16.240***$ (0.003)	1.42 (0.583)
<b>Presentation</b>	0.1290* (0.058)	$0.1102***$ (0.001)	0.8666 (0.662)	0.8046 (0.788)	2.02 (0.310)
<b>Diagram Task</b>	5.38* (0.071)	5.11* (0.075)	0.9186 (0.785)	$2.714**$ (0.049)	1.21 (0.710)
<b>Constant</b>	1.83 (0.961)	0.0284 (0.894)	0.0550 (0.766)	$3.00 e^7$ (0.226)	583.06 (0.595)
<b>Wald <math>\chi^2</math></b>	25.71*** (0.0072)	27.56*** (0.0038)	18.69* (0.0669)	27.54*** (0.0038)	27.43*** (0.0039)
<b>Pseudo R<sup>2</sup></b>	0.4339	0.5080	0.2532	0.6917	0.4956

Note: We report odds ratios from logistic regressions on dummy variables that indicate rational group decisions according to SEUT, ambiguity preferences measured over all gambles, and ambiguity preferences measured over gambles with an associated objective probability of 0.5. The p-values are reported in parentheses. \*\*\*, \*\*, and \* indicate statistically significant deviations from 1. Ambiguity preference and preference shift parameters are calculated according to the applied rationality criterion in each model. “Average Individual”, “Best Individual”, and “Worst Individual” refer to the way the groups’ skill measures are incorporated in the regression models.

First, note that the odds ratios for the treatment dummies are all below one, even though the difference is only significant for two of the models. Thus, group decisions are less likely to be

consistent with SEUT or group ambiguity preferences if the subjects have taken the individual treatment first. We assume that it is more difficult for the group members to agree on consistent choice patterns in this case, because they leave the individual stage of the experiment with well-founded and potentially very different preferences and opinions.

Comparing the models for the two different rationality criteria under ambiguity ( $\gamma_j$  over all gambles and over gambles with  $p^{obj} = 0.5$ ), using the scores of the best individual yields higher explanatory power as measured by Pseudo  $R^2$  than the scores of the worst individual or the average. Furthermore, using the ambiguity preference parameters derived from the gambles with an objective probability of fifty-fifty generally yields the better overall fit. Here, higher values for the preference parameter  $\gamma_j$ , i.e. less ambiguity aversion or even more ambiguity seeking, significantly reduce rationality, which we have already seen in the analysis on the individual level. Preference shifts seem to work into the same direction, but the result is somewhat unstable as it only occurs when we use the scores of the best individual for the whole group, even though the values for  $\gamma_j$  and  $\Delta\gamma_j$  are the same in both variations. Consistency with SEUT for groups is significantly influenced by group shifts in ambiguity preferences, which is consistent with the shift patterns that we discussed in section 2.6.1.

Interestingly, contrary to individual decisions, group decisions in the Ellsberg Paradox are significantly influenced by a number of the personal skill variables, especially when we use the best score of all the team members for a specific skill and attribute it to the group, but also in other specifications. Higher scores in goal orientation, the group discussion, and the diagram task significantly increase rationality of the group decision. However, even though all the properties that we measure have a clearly positive connotation and could therefore be expected to have positive influences on decision making and problem solving capabilities, some others have a significantly negative influence on group rationality in the Ellsberg Paradox. Surprisingly, higher presentation scores and better peer evaluations in analytical competence, persuasiveness, and in leadership significantly reduce rationality in different specifications of the model. Finally, the evidence on grades in secondary school is mixed across different specifications.

For the Zeckhauser Paradox, all three regression models fail to provide significant explanatory power in the Russian Roulette version. This was also the case in individual decisions. However, in the Dice Game version, using the average skills and the scores of the

worst team member yields overall significant results. The following table summarizes these two models.

Table 2.20: Determinants of Consistency of Group Decisions in the Dice Game version of the Zeckhauser Paradox (N = 36)

	Average individual	Worst individual
<b>Treatment Dummy (1 = individual treatment first)</b>	0.3608 (0.480)	0.8473 (0.876)
$\lambda_j$	86.08 (0.552)	0.9534 (0.995)
$\Delta\lambda_j$	2.33 (0.893)	504.39 (0.238)
<b>Analytical Competence</b>	0.2150 (0.372)	0.1643*** (0.005)
<b>Goal Orientation</b>	0.5103 (0.554)	3.82 (0.123)
<b>Persuasiveness</b>	0.0198** (0.014)	0.0411** (0.022)
<b>Leadership</b>	6.6290 (0.145)	8.2663*** (0.008)
<b>Final Grade Secondary School</b>	4.84 (0.589)	18.0154** (0.045)
<b>Group Discussion</b>	0.8856 (0.867)	1.70 (0.489)
<b>Presentation</b>	0.4837 (0.310)	1.19 (0.699)
<b>Diagram Task</b>	0.4872 (0.238)	0.7725 (0.644)
<b>Constant</b>	11790.83 (0.368)	0.0038 (0.470)
<b>Wald <math>\chi^2</math></b>	22.50 (0.0207)**	18.64* (0.0679)
<b>Pseudo R<sup>2</sup></b>	0.3965	0.4440

Note: We report odds ratios from logistic regressions on dummy variables that indicate rational group decisions according to EUT. The p-values are reported in parentheses. \*\*\*, \*\*, and \* indicate statistically significant deviations from 1. “Average Individual”, “Best Individual”, and “Worst Individual” refer to the way the groups’ skill measures are incorporated in the regression models.

First, note that, like in the individual case, risk preferences and risk shifts do not significantly influence rationality of the group decision in the Dice Game. Persuasiveness, as seen before, has a significant negative influence in group decisions, too. This holds for both models, i.e. when using the average scores and the worst scores that are available. All other skill variables

do not produce odds ratios significantly different from one when we look at averages. The higher the analytical competence and the better the final grade from secondary school of the worst individual, the more the group decision tends to be irrational. Stronger leadership skills of the worst individual, in turn, foster group rationality in the Dice Game.

Overall, we attest that the personal skills of the team members are important for the outcomes of the group decisions, but there is more research needed to understand the underlying processes. For example it could be fruitful to conduct in-depth interviews with the test persons right after they take part in the experiment in order to learn more about their thoughts on how their teams have come to their decisions and whether specific skills or characteristics are perceived as helpful, relevant, or disadvantageous. Interestingly, all the skill variables that we use have a significant influence on group rationality in at least one of the models that we present. On the other hand, none of the variables is significant in all the models, and we also must not forget that there are two more variations, namely SEUT in the Ellsberg Paradox and the Russian Roulette version of the Zeckhauser Paradox, for which our models totally fail to explain group rationality. Thus, we should not take our findings as fundamental principles but we have to keep in mind that the specific context of the decision in question matters.

Analytical competence has a significant negative influence on group rationality in both paradoxes, but not in all the models that we test. Also, the result is somewhat surprising and needs to be treated carefully, as the score of the diagram task, which is also designed to test for analytical skills, works into the other direction in the Ellsberg Paradox. Goal Orientation positively influences rationality in both paradoxes, but also not in all models. Persuasiveness only plays a role in the Zeckhauser Paradox, where it decreases the odds to come to a rational decision. The score for the presentation from the assessment center, on the other hand, only matters in the Ellsberg Paradox, where it supports rationality – but only in one of the four models that we test here. The evidence for the peer evaluation in leadership and the final grade from secondary school is mixed: in the Ellsberg Paradox, a higher leadership score reduces rationality while a better grade supports consistency – and in the Zeckhauser Paradox, the interdependencies are exactly the other way around.

Which approach should we choose to describe the skills of a team? In the Ellsberg Paradox, the models in which we attribute the best available score to the whole team yield the highest fit in terms of the Pseudo- $R^2$ . However, in the Zeckhauser Paradox, this approach does not produce any significant results at all. Instead, the best performing model uses the scores of the worst individual. The regressions where we use the average scores generally have the lowest

fit. From these different findings for the two paradoxes, we can conclude that under ambiguity the rationality of group decisions is driven by extreme individual occurrences of personal and social skills: the higher the score of the best individual, the more do they influence the team's rationality, while the correspondent scores of the other team members are irrelevant. In the Zeckhauser Paradox, i.e. when dealing with risk, it is crucial that the lowest score among the group members is high enough in order to attest an influence of the respective skill variable on the rationality of the group decision. It is not sufficient that a single individual stands out in terms of a specific skill, but all team members need to have relatively high scores. The worst individual serves as the lower boundary, and significant influences can only be seen when this boundary is pushed upwards, no matter how far the other team members' scores are above. However, we would like to point out once more that high parameter values, for either one individual or the whole team, are not per se desirable, as some of the skills have been shown to be detrimental for rational outcomes.

#### 2.6.6 Preference Shifts and Personal and Social Skills

We conclude this section by turning to research question d) – we will discuss the interrelation between risk and ambiguity preference shifts and the personal and social skill measures that we have at hand. On the group level, when comparing actual group decisions in the BDM lotteries and gambles with the hypothetical group decisions derived from the individual decisions of the group members, we do observe a conversion towards risk neutrality, while there are no significant changes in ambiguity preferences. This difference ( $\Delta\lambda_j$  or  $\Delta\gamma_j$ ) between the actual and the hypothetical (i.e. the mean of the group members) preference parameters  $\lambda$  and  $\gamma$  describes the overall preference shift caused by group formation. However, by simply averaging over the three individual preference parameters, this calculation method neglects the fact that for each team member, the difference between their personal preference parameter  $\lambda_i$  or  $\gamma_i$  and the common parameter  $\lambda_j$  or  $\gamma_j$  is unequal. Even in absence of a group preference shift, the group decision usually differs from the individual decisions. Therefore, we define an individual preference shift as the difference between the group preference parameter and the individual preference parameter.

$$\Delta\lambda_i = \lambda_j - \lambda_i$$

$$\Delta\gamma_i = \gamma_j - \gamma_i$$

To compare the magnitude of the individual shifts within each team, we use the absolute differences. The individual with the largest individual shift (regardless of the direction) is the one whose personal preference is least reflected in the team decision. To analyze whether there are structural differences in our miscellaneous personal skill variables that depend on preference shifting behavior, we proceed as follows. We rank the individual shifts from lowest to highest for each group. Afterwards, we construct three subsamples using these shift ranks. For each subsample, we calculate the mean ranks of all individuals for the personal and social skill measures from the peer evaluation and the assessment center and test for a significant difference between the three subsamples with a Kruskal Wallis H-Test. A higher mean rank indicates a higher numerical value for the respective skill variable, which translates into a higher skill level, with the only exception being the final grade from secondary school, where better test persons have lower numerical values. Besides the overall difference between all mean ranks, we also make pairwise comparisons by applying Mann Whitney U-Tests. For the individual risk preference shifts, the results are as follows. The grade from secondary school is the only variable where the H-Test indicates a statistical significance between all three samples. The persons whose preferences have shifted the least are the ones with the best grades and those whose preferences are changed the most have the worst grades. The pairwise comparison between the lowest and the highest shift subsample also yields a statistically significant difference.<sup>37</sup> The point scores for the presentation earned in the assessment center also differ between the risk shifting subsamples. The individuals in the lowest shifting tercile have significantly higher scores than the ones in the middle tercile, but the ones who shift most exhibit a mean rank between the other two subsamples which is not significantly different from theirs. For all other skill variables, we do not find any significant differences between the risk shifting subsamples. Therefore we can conclude that skill factors play at most a subordinate role in risk preference shifts.

---

<sup>37</sup> In Table 2.21, the results of the pairwise comparisons are indicated by lower case characters in the subscripts of the subsample mean ranks. If two mean ranks are marked with at least one identical subscript, there is no statistically significant difference between the two. On the other hand, if they do not have at least one letter in common, the test indicates a significant difference. In the case of the final grade from secondary school in Table 2.21, the mean rank of 65.31 (marked with letter “b” in the subscript) is significantly different from both, 44.53 and 50.91, because neither of them has a “b” in the subscript. However, there is no significant difference between 44.53 and 50.91 as both have the letter “a” in their subscript. We use this notation several times throughout all chapters of the dissertation.

Table 2.21: Personal Skills and Individual Risk Preference Shifts

Individual risk preference shift rank (as measured by $\Delta\lambda_i$ )		N	Mean Rank
Analytical competence	1	36	57.74 <sub>a</sub>
	2	36	56.40 <sub>a</sub>
	3	36	49.36 <sub>a</sub>
Goal-orientation	1	36	58.36 <sub>a</sub>
	2	36	58.07 <sub>a</sub>
	3	36	47.07 <sub>a</sub>
Persuasiveness	1	36	55.61 <sub>a</sub>
	2	36	58.38 <sub>a</sub>
	3	36	49.51 <sub>a</sub>
Leadership	1	36	55.29 <sub>a</sub>
	2	36	56.75 <sub>a</sub>
	3	36	51.46 <sub>a</sub>
Final grade secondary school**	1	36	44.53 <sub>a</sub>
	2	35	50.91 <sub>a</sub>
	3	35	65.31 <sub>b</sub>
Diagram task	1	36	60.29 <sub>a</sub>
	2	36	54.36 <sub>a</sub>
	3	36	48.85 <sub>a</sub>
Group discussion	1	36	50.42 <sub>a</sub>
	2	36	52.17 <sub>a</sub>
	3	36	60.92 <sub>a</sub>
Presentation	1	36	61.64 <sub>a</sub>
	2	36	46.53 <sub>b,c</sub>
	3	36	55.33 <sub>a,c</sub>

Note: Mean ranks reported in the last column are computed in a Kruskal Wallis H-Test, with low ranks indicating smaller numerical values in the respective skill variables. \*\*\*, \*\*, and \* indicate a statistically significant difference of these mean ranks between the three samples with the lowest (1), medium (2) and highest (3) risk preference shift (as measured by absolute values of  $\Delta\lambda_i$ ) on the 99%, 95%, and 90% confidence level, respectively. Mean ranks *not* sharing an identical subscript are statistically significantly different from each other on at least the 90% confidence level, as indicated by pairwise comparisons using Mann Whitney U-Tests.

When we look at the preference shifts observable in decisions under ambiguity, the picture is different. Pairwise Mann Whitney U-Tests reveal significant differences between the lowest and the highest shifters, with the least shifting individuals having the better scores for goal-orientation, persuasiveness, the diagram task, and the presentation. The scores in the middle sample are either not significantly different from one or from both of the other two subsamples. A somewhat different result occurs for the group discussion, which is also the only skill variable that is positively tested for differences between all three samples in the H-Test. Here, the test persons who shift most have the highest scores, followed by the ones who shift least, and the individuals in the middle tercile have the lowest point scores. The reason for this could be that the persons whose individual ambiguity preferences are least reflected in the resulting group decision, i.e. who are in the highest shifting subsamples, are the ones who are most willing to adjust their own position in the process of finding a decision in the group, who are most open for their teammates' arguments and who tend to accept quickly that they might have a majority against them without insisting too much. Even though this kind of behavior diminishes chances to arrive at the individual's personal preference with the whole group, it is still beneficial for earning high scores in the assessment center's group discussion. Contrary to individual risk shifting, the final grade from secondary school does not play a role in ambiguity preferences, and so do the peer evaluations in analytical competence and leadership.



Table 2.22: Personal Skills and Individual Ambiguity Preference Shifts

Individual ambiguity preference shift rank (as measured by $\Delta\gamma_i$ for gambles with $p^{obj} = 0.5$ )		N	Mean Rank
Analytical competence	1	36	54.26 <sub>a</sub>
	2	36	55.80 <sub>a</sub>
	3	36	53.46 <sub>a</sub>
Goal-orientation	1	36	60.17 <sub>a</sub>
	2	36	56.33 <sub>a,b</sub>
	3	36	46.51 <sub>b</sub>
Persuasiveness	1	36	59.83 <sub>a</sub>
	2	36	55.77 <sub>a,b</sub>
	3	36	47.44 <sub>b</sub>
Leadership	1	36	57.72 <sub>a</sub>
	2	36	58.40 <sub>a</sub>
	3	36	47.10 <sub>a</sub>
Final grade secondary school	1	36	50.54 <sub>a</sub>
	2	35	52.97 <sub>a</sub>
	3	35	57.34 <sub>a</sub>
Diagram task	1	36	58.76 <sub>a</sub>
	2	36	58.79 <sub>a</sub>
	3	36	45.59 <sub>b</sub>
Group discussion**	1	36	55.50 <sub>a,c</sub>
	2	36	44.83 <sub>a</sub>
	3	36	63.09 <sub>b,c</sub>
Presentation	1	36	58.25 <sub>a</sub>
	2	36	54.06 <sub>a,b</sub>
	3	36	50.87 <sub>b</sub>

Note: Mean ranks reported in the last column are computed in a Kruskal Wallis H-Test, with low ranks indicating smaller numerical values in the respective skill variables. \*\*\*, \*\*, and \* indicate a statistically significant difference of these mean ranks between the three samples with the lowest (1), medium (2) and highest (3) ambiguity preference shift (as measured by absolute values of  $\Delta\gamma_i$  for gambles with  $p^{obj} = 0.5$ ) on the 99%, 95%, and 90% confidence level, respectively. Mean ranks *not* sharing an identical subscript are statistically significantly different from each other on at least the 90% confidence level, as indicated by pairwise comparisons using Mann Whitney U-Tests.

## 2.7 Discussion

In this chapter of the dissertation we show that rational decision making in some versions of the paradoxes that we present, as well as preference formation in groups – especially in ambiguous situations, under risk only to a minor degree – can partly be explained by personal and social skill variables. In this concluding section, we summarize our results and we discuss some proposals for potential advancement of the literature on group dynamics in decision making.

First of all, group choices are not more rational than individual decisions. Overall, we observe that inconsistent choice behavior is very prevalent in all variations of the paradoxes that we test, except for the Allais Paradox. This is true for individual choices and group choices. We neither find significant differences in rationality between individuals and groups in the Common Ratio and the Common Consequences version of the Allais Paradox, nor in the Dice Game version and the Russian Roulette version of the Zeckhauser Paradox. In the Ellsberg Paradox, when we use the ambiguity preference derived from the full set of BDM gambles as the criterion, group decisions are even significantly worse. However, there are no significant differences in the Ellsberg Paradox if we use SEUT or the ambiguity preferences derived from the subset of gambles that have an associated objective probability of 0.5 as the criteria for rationality.

On an individual basis, personal and social skills do not determine rational choice behavior. In the Allais Paradox, there is practically no variance in our observations so that we cannot use the data for the analyses. In the Ellsberg Paradox, ambiguity aversion is the only personal property that fosters rationality – if we apply this specific preference attribute as the criterion for choices to be classified as consistent. Otherwise – under SEUT – none of our measures has any explanatory power. In the Zeckhauser Paradox, neither risk preferences nor any of the personal or social skill variables have a significant influence on rationality. Only in the Dice Game version, male subjects are significantly less likely to make consistent valuations than females.

Group formation causes changes in risk preferences, but not in ambiguity preferences. For risk preferences, we do not see an overall shift in preferences, i.e. both individuals and groups exhibit the same moderate degree of risk aversion on average. However, groups that consist of highly risk averse subjects tend to be less risk averse than expected and groups that have highly risk seeking members are less risk seeking than we would predict from the individual

preferences. For individuals and groups the elicited ambiguity preferences are practically identical on average. In contrast to risk preferences, there is no conversion towards the mean. Even if there is no overall preference shift observed for a specific group, i.e. if the group preference parameter equals the average of the individual group members' preference parameters, the group decision does usually not reflect the individual preferences of each single group member. In fact, this would only be the case, if the preference parameters were identical for all members. The degree to which an individual's risk preference is reflected in the group preference parameter is related to two of the skill measures that we use. The test person with the best final grade in secondary school among the group members has to adjust her risk preference the least and persons with weaker grades need to adjust more. A similar result is obtained when looking at the presentation score. For individual shifts in ambiguity preferences, we find significant differences in even more skill measures, namely in goal-orientation, persuasiveness, the diagram task, the group discussion, and the presentation scores. Thus, the adjustment of individual preferences in the group decision process is strongly related to personal and social skills for ambiguous choices, and to a lesser degree for risky choices.

For group decisions, we find that consistency in the Ellsberg Paradox and the Zeckhauser Paradox can be partly explained by personal and social skills, but more research is needed to understand the dynamics. We have to admit that the skill variables that we use probably suffer from several shortcomings. The peer evaluations are potentially imprecise, since the test persons may have different conceptions of the attributes in question. For example, "leadership" is a very versatile term. On the other hand, the ability of the test persons to recognize and evaluate personal attributes of their peers may be limited. For example, we observe that all variables stemming from the peer evaluations are quite highly positively correlated – the Spearman correlation coefficients range from 0.52 to 0.77 and all correlations among these variables are significantly different from zero. For the multivariate regressions where we use these variables, the correlations are still at an acceptable level, i.e. our results do not suffer from multicollinearity, but they demonstrate the conceptual weakness of the peer evaluation. The scores obtained from the assessment center data and the test persons' final grades of secondary school are surely more objective. The jury members who are involved in the selection process for the study program run through a detailed training where they are prepared thoroughly for their function. There is a catalog of requirements for each part of the assessment center that defines the criteria by which the jury members have to award their

point scores to the applicants. We also observe that the variables from the assessment center are uncorrelated amongst each other. Only the grade from secondary school is weakly but significantly negatively correlated with the point scores for the presentation (Spearman correlation coefficient -0.26) and the diagram task (-0.17), i.e. with better grades (smaller figures represent higher achievements in German school grades), the scores for these tasks tend to be higher. The clear disadvantage of the assessment center variables, however, is at the same time the strength of the peer evaluations: while the peer evaluations are made immediately after the experiment and the test persons are asked to refer their judgment to the decision making process in the group treatment, the assessment center has taken place well before the experiment is run – depending on the test person the elapsed time ranges between a couple of months up to two years.

This dilemma can probably be overcome by using established psychological tests to measure personal traits and skills. These tests should be done immediately before or after the actual experiment. However, measuring psychological variables is complex and time-consuming (for example, the assessment center in the university's selection process lasts for a whole day), and with an average time of about two hours per test person, our experiment has already been very extensive. Therefore we suggest that, instead of analyzing the effects of multiple skill variables, as we do in this study, future experiments should focus on single traits.

Our results also suggest that, even though there are indeed some significant influences that we detect, the role of personal and social skills in group decision making is very context sensitive. This stresses the fact that much more research is necessary to develop a more complete understanding of the matter. We also see that the interdependencies between skills and decisions and preferences do not follow straightforward patterns like, for example “the higher the skill level, the more likely is the decision to be rational”. This is true for some skills in some specific decisions, but by no means a general principle. For future research, this brings up the question of how exactly a personal attribute exerts influence in a decision making process. In our data, we can only observe correlations, but not functionalities, i.e. causation. It might therefore be fruitful to record the decision making processes of the teams by, for example, letting them communicate via a chat protocol only to gain more qualitative insights. Also, ex post interviews with the individual team members could be viable for this purpose. However, both extensions of our experimental setup potentially suffer from an additional degree of subjectivity of the experimenter in interpreting the additional data.

Finally, even though we do not find that teams are more rational than individuals, our findings suggest that in principle it is possible to foster higher degrees of consistency in team decisions by carefully selecting the team members by their individual skills. However, within the limits of our experiment, and at the current level of knowledge, we have to refrain from giving specific advice on how this needs to be done.

## **Appendix to Chapter 2**

### **Appendix 2.A / Individual & Group Treatment, Introduction Screen (Translation)**

Before the start of the experiment the participants were informed that participation in the experiment is voluntary and their data will only be reported in an anonymous form. After signing a consent form, the experiment started. In the individual treatment, subjects were shown this introduction screen. [Instructions for the group treatment that differ from the individual treatment are written in square brackets after the respective text passage.]

“Welcome and thank you very much for your participation in this experiment!

This experiment is conducted by the chair of Empirical Capital Market Research at WHU – Otto Beisheim School of Management and the collected data will be used in the dissertation of Maximilian Trossbach.

First we would like to inform you that your data will be treated confidentially. The identification by your email address is only necessary to match your data from different parts of the experiment and to inform you about your profit that you made in the experiment. Your name will not be published and it will not be possible to draw any conclusions to you personally on the basis of published data.

The experiment consists of three parts. Please answer this first part on your own [Please answer this second part together in the team that you have been assigned to] and without use of any auxiliary devices. Please shut down all other active programs on your computer. We are interested in your personal and independent decisions [We are interested in the decisions that you make as a team]; therefore the use of any auxiliary material or communication with other participants [teams] would produce undesired results.

In the following you are presented a number of gambles (lotteries) that differ in their potential gains and losses and the probabilities with which these outcomes occur. You will be first asked to evaluate different lotteries. Imagine that you can sell your lottery ticket, i.e. the opportunity to take part in the lottery, to someone else instead of using it for yourself. In some other tasks of this questionnaire you are asked to choose among different lotteries the one you would like to play best. There are no objectively right or wrong answers to any of the following questions. [*addition for group treatment only*: It is necessary that you agree as a team on a common, unanimous valuation or choice! In case you cannot agree, please leave out

the respective answer. You will then receive the pre-specified standard payoff for this question.] You will receive more precise instructions during the course of the experiment.

At the end of the experiment, three of the lotteries will be chosen randomly and will actually be played. You will then receive the payoffs corresponding to your choices. Thus, you should pay attention to answer truthfully in every single decision. This also contributes to the optimal usability of your data.

Throughout the whole experiment we will use the virtual currency MU (= monetary units). All gains and losses are denoted in MU and you will be asked to make your evaluations in MU. At the end of the experiment, you will receive your profit in Euros. The conversion rate is the same for all participants and in all parts of the experiment and is not disclosed in advance.

Please start with the first part of the experiment now. It will take about 30 minutes.

Please do not use the “back” and “forward” keys of your browser. Use the navigation buttons in the questionnaire instead.”

## **Appendix 2.B / Instruction: Individual & Group Treatment, Risk and Ambiguity Preference Elicitation (Translation)**

[Instructions for the group treatment that differ from the individual treatment are written in square brackets after the respective text passage.]

“In the following we present you a number of lotteries. Imagine that you have a ticket to participate in each of these lotteries that you can either use for yourself or that you can sell to someone else. The lotteries have a certain probability (either 25%, 50%, 75%, or *unknown*) to win amount X. With the complementary probability you can win amount Y. [*addition for group treatment only*: The amounts indicate the total profits for your team. You are free to agree on any distribution of the total profit among your team members.]

All lotteries are described exactly like this example:

In this lottery you can win the amount X with a 50% probability. Otherwise you win Y.

**X: 10,000 MU [30,000 MU]**

**Y: 0 MU**

**Probability for X: 50%**

We ask you to indicate the price in MU for each lottery, for which you are only just willing to sell the lottery (= selling price). This is the price that has at least to be offered to you to sell your lottery ticket instead of playing yourself. Thus, please determine the price for which selling your lottery ticket and playing the lottery is equally desirable for you: at any higher price you would certainly like better to sell your ticket instead of playing and for any lower price you would like better to play yourself instead of selling.

At the end of the experiment, an offer price that is between X and Y (between the highest and the lowest possible outcome of each lottery) will be determined randomly for each lottery. If the offer price is higher than the selling price or if offer price and selling price are identical, you sell the lottery for your indicated selling price. If the offer price is below your selling price, you play the lottery.

It is reasonable to indicate your selling price truthfully: in case you report a selling price that is higher than your actual selling price, it can happen that you receive an offer price that falls into the gap between your true selling price and your stated selling price. In this case, you would rather sell the lottery ticket, but now you have to play the lottery.



Conversely, with a pretended selling price that is below your true selling price, it can happen that you have to sell a lottery ticket at a price that is below your valuation of that lottery. You can also prevent this by stating your true valuation as your selling price.

Please answer the following questions chronologically by stating your minimum demanded selling price for the lottery ticket and clicking the “next” button afterwards. You cannot revise your answers later on. Therefore please determine your selling prices carefully. If you have any questions, please ask the experimenter.

Before we start, we illustrate the procedure step by step in a practice round.

Reconsider the previous example:

**X: 10,000 MU [30,000 MU]**

**Y: 0 MU**

**Probability for X: 50%**

You have a ticket for this lottery. Now think about how much this lottery is worth for you [for your team]. Therefore, it may be helpful to calculate the expected profit. In case of this example, it is 5,000 MU [15,000 MU]. But, as you can only win 10,000 MU [30,000 MU] or 0 MU when playing the lottery, the profit is *uncertain*. When evaluating the lottery you should decide [*addition for group treatment only*: together and unanimously in your team] which *certain* amount must at least be offered to you [your team] that you sell the lottery ticket and cannot play the lottery yourself anymore.

[*addition for group treatment only*: Please do always (in this example in all following decisions during this part of the experiment) state the total amount that you demand as the minimum selling price for your whole team. The selling price will not be multiplied by the number of members in your team. This means that if your team, for example, has three members and you agree on a selling price of 12,000 MU (the lottery pays you either 0 MU or 30,000 MU), then please do also state this price and not 4,000 MU! It is up to you how you distribute the selling price or the lottery payoffs among your team members.]

My [Our] selling price for this lottery ticket in MU is \_\_\_\_

Please enter a whole number (e.g. 1250 or 8300, not 4000.7) [(e.g. 12,000 or 25,000, not 4000.7)] in the field.

Good! You indicated for which amount you would be willing to sell your lottery ticket. In this experiment we are interested in the minimum amount that require for selling the ticket. If you

were offered less than this amount, you would certainly prefer to keep your ticket and play the lottery.

You have indicated that you would like to receive at least \_\_\_\_ MU for selling your lottery ticket. Please rethink again carefully: would you also accept less than this? Or would you really strictly prefer to play the lottery at any lower selling price?

Maybe you are thinking now that it would be unwise to admit which is in fact the lowest acceptable selling price for you. After all, receiving more is better than less. However, you have to consider that the probability that you can actually sell your ticket for a certain amount (your stated selling price) declines when you set the price higher. The sale will only take place if the offer that you receive after your decision is at least as high as your selling price. Otherwise you will play the lottery and cannot sell the ticket.

The offer price that you receive is determined randomly and can take on any value between X and Y, in this case between 0 MU and 10,000 MU [30,000 MU]. A very high selling price of e.g. 9000 MU [27,000 MU] means that you can only sell the ticket at this price with a 10% chance and conversely, you will play the lottery with a 90% chance. A very low selling price of e.g. 1000 MU [3000 MU] means that you will sell the ticket with a 90% chance at this price and you will play the lottery with only a 10% chance.

Please reconsider what your selling price for this lottery really is. If the offer price is subsequently below your selling price you play the lottery – and this is also in line with your preferences, as you would otherwise receive less than the lottery is worth for you. If the offer price is higher or identical with your selling price, then you receive your selling price and do not play the lottery.

When you finally made your decision, please click “next”. The first lottery that you are asked to evaluate is identical with the example above. If you have any further questions, please ask the experimenter before clicking “next”.

## Appendix 2.C / Individual & Group Treatment, Risk and Ambiguity Preference Elicitation (Translation)

A much shorter version of the instructions in Appendix 2.B is presented to the test persons with each of the following lotteries. The lotteries are adopted from Becker et al. (1964). We transformed the nominal payoff range to values between 0 and 10,000 MU (group treatment: 0 and 30,000 MU to keep the range of per capita payoffs constant). Furthermore, we extend their set of lotteries by adding ambiguous gambles with unknown probabilities (for example, lottery 3A is the corresponding ambiguous gamble for lottery 3). Some payoffs in the schedule below are fixed, while others are identical with the selling price that participant  $i$  or group  $j$  has reported for a previous lottery  $k$  (indicated by the variables  $SP_{i,k}^{risk}, SP_{i,k}^{amb}, SP_{j,k}^{risk}, SP_{j,k}^{amb}$ ), with  $k = 1, \dots, 24$ .  $X_k$  and  $Y_k$  are the per capita payoffs of the binary lotteries and gambles,  $\text{prob}(X_k)$  is the probability that lottery  $k$  pays amount  $X_k$  and  $U(SP_{i/j,k}^{risk})$  indicates the utility that subject  $i$  or group  $j$  assigns to both, lottery  $k$  and the reported certainty equivalent  $SP_{i/j,k}^{risk}$ . For ambiguous gambles, utility values cannot be calculated due to the unknown probabilities.

$k$	$X_k$	$Y_k$	$\text{prob}(X_k)$	$SP_{i/j,k}^{risk/amb}$	$U(SP_{i/j,k}^{risk})$
1	10,000	0	0.5	$SP_{i/j,1}^{risk}$	5,000
2	10,000	0	0.75	$SP_{i/j,2}^{risk}$	7,500
3	10,000	0	0.25	$SP_{i/j,3}^{risk}$	2,500
3A	10,000	0	unknown	$SP_{i/j,3}^{amb}$	n/a
4	10,000	$SP_{i/j,1}^{risk}$	0.5	$SP_{i/j,4}^{risk}$	7,500
4A	10,000	$SP_{i/j,1}^{risk}$	unknown	$SP_{i/j,4}^{amb}$	n/a
5	$SP_{i/j,1}^{risk}$	0	0.5	$SP_{i/j,5}^{risk}$	2,500
5A	$SP_{i/j,1}^{risk}$	0	unknown	$SP_{i/j,5}^{amb}$	n/a
6	$SP_{i/j,3}^{risk}$	$SP_{i/j,2}^{risk}$	0.75	$SP_{i/j,6}^{risk}$	3,750
6A	$SP_{i/j,3}^{risk}$	$SP_{i/j,2}^{risk}$	unknown	$SP_{i/j,6}^{amb}$	n/a
7	$SP_{i/j,3}^{risk}$	$SP_{i/j,2}^{risk}$	0.25	$SP_{i/j,7}^{risk}$	6,250
7A	$SP_{i/j,3}^{risk}$	$SP_{i/j,2}^{risk}$	unknown	$SP_{i/j,7}^{amb}$	n/a

$k$	$X_k$	$Y_k$	$prob(X_k)$	$SP_{i/j,k}^{risk/amb}$	$U(SP_{i/j,k}^{risk})$
8	$SP_{i/j,3}^{risk}$	10,000	0.25	$SP_{i/j,8}^{risk}$	8,125
8A	$SP_{i/j,3}^{risk}$	10,000	unknown	$SP_{i/j,8}^{amb}$	n/a
9	10,000	$SP_{i/j,5}^{risk}$	0.5	$SP_{i/j,9}^{risk}$	6,250
9A	10,000	$SP_{i/j,5}^{risk}$	unknown	$SP_{i/j,9}^{amb}$	n/a
10	$SP_{i/j,1}^{risk}$	$SP_{i/j,5}^{risk}$	0.5	$SP_{i/j,10}^{risk}$	3,750
10A	$SP_{i/j,1}^{risk}$	$SP_{i/j,5}^{risk}$	unknown	$SP_{i/j,10}^{amb}$	n/a
11	$SP_{i/j,3}^{risk}$	0	0.25	$SP_{i/j,11}^{risk}$	625
11A	$SP_{i/j,3}^{risk}$	0	unknown	$SP_{i/j,11}^{amb}$	n/a
12	10,000	$SP_{i/j,2}^{risk}$	0.75	$SP_{i/j,12}^{risk}$	9,375
12A	10,000	$SP_{i/j,2}^{risk}$	unknown	$SP_{i/j,12}^{amb}$	n/a
13	10,000	$SP_{i/j,4}^{risk}$	0.5	$SP_{i/j,13}^{risk}$	8,750
13A	10,000	$SP_{i/j,4}^{risk}$	unknown	$SP_{i/j,13}^{amb}$	n/a
14	$SP_{i/j,8}^{risk}$	$SP_{i/j,7}^{risk}$	0.75	$SP_{i/j,14}^{risk}$	7,656.25
14A	$SP_{i/j,8}^{risk}$	$SP_{i/j,7}^{risk}$	unknown	$SP_{i/j,14}^{amb}$	n/a
15	$SP_{i/j,4}^{risk}$	$SP_{i/j,5}^{risk}$	0.5	$SP_{i/j,15}^{risk}$	5,000
15A	$SP_{i/j,4}^{risk}$	$SP_{i/j,5}^{risk}$	unknown	$SP_{i/j,15}^{amb}$	n/a
16	10,000	$SP_{i/j,9}^{risk}$	0.5	$SP_{i/j,16}^{risk}$	8,125
16A	10,000	$SP_{i/j,9}^{risk}$	unknown	$SP_{i/j,16}^{amb}$	n/a
17	$SP_{i/j,7}^{risk}$	$SP_{i/j,6}^{risk}$	0.75	$SP_{i/j,17}^{risk}$	5,625
17A	$SP_{i/j,7}^{risk}$	$SP_{i/j,6}^{risk}$	unknown	$SP_{i/j,17}^{amb}$	n/a
18	$SP_{i/j,13}^{risk}$	$SP_{i/j,9}^{risk}$	0.5	$SP_{i/j,18}^{risk}$	7,500
18A	$SP_{i/j,13}^{risk}$	$SP_{i/j,9}^{risk}$	unknown	$SP_{i/j,18}^{amb}$	n/a
19	$SP_{i/j,8}^{risk}$	$SP_{i/j,11}^{risk}$	0.75	$SP_{i/j,19}^{risk}$	6,250
19A	$SP_{i/j,8}^{risk}$	$SP_{i/j,11}^{risk}$	unknown	$SP_{i/j,19}^{amb}$	n/a
20	$SP_{i/j,13}^{risk}$	0	0.5	$SP_{i/j,20}^{risk}$	4,375
20A	$SP_{i/j,13}^{risk}$	0	unknown	$SP_{i/j,20}^{amb}$	n/a

$k$	$X_k$	$Y_k$	$prob(X_k)$	$SP_{i/j,k}^{risk/amb}$	$U(SP_{i/j,k}^{risk})$
21	$SP_{i/j,4}^{risk}$	$SP_{i/j,1}^{risk}$	0.5	$SP_{i/j,21}^{risk}$	6,250
21A	$SP_{i/j,4}^{risk}$	$SP_{i/j,1}^{risk}$	unknown	$SP_{i/j,21}^{amb}$	n/a
22	$SP_{i/j,8}^{risk}$	$SP_{i/j,11}^{risk}$	0.25	$SP_{i/j,22}^{risk}$	2,500
22A	$SP_{i/j,8}^{risk}$	$SP_{i/j,11}^{risk}$	unknown	$SP_{i/j,22}^{amb}$	n/a
23	$SP_{i/j,4}^{risk}$	0	0.5	$SP_{i/j,23}^{risk}$	3,750
23A	$SP_{i/j,4}^{risk}$	0	unknown	$SP_{i/j,23}^{amb}$	n/a
24	10,000	$SP_{i/j,11}^{risk}$	0.75	$SP_{i/j,24}^{risk}$	7,656.25
24A	10,000	$SP_{i/j,11}^{risk}$	unknown	$SP_{i/j,24}^{amb}$	n/a

## Appendix 2.D / Individual & Group Treatment, Choice Paradoxes (Translation)

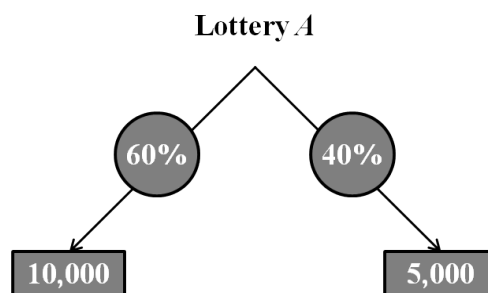
In this part of the Appendix we provide translations of all versions of the choice paradoxes that we tested in the experiment. Each question or decision was described in a separate window and in a non-randomized order. Associated questions were always displayed directly in succession. [Instructions for the group treatment that differ from the individual treatment are written in square brackets after the respective text passage.]

### Allais Paradox (common consequences), between lotteries 7A and 8 (see Appendix 2.C)

In the following, we present to you two pairs of lotteries between which you have to decide. [*addition for group treatment only:* You shall make an unanimous decision as a team, i.e. each team member has to agree to the choice]. Lotteries are described verbally and graphically. The probabilities of the outcomes are stated in percentages and the value of the outcomes is stated in “monetary units” MU [*addition for group treatment only:* and are to be understood as total payoffs for your team. You may agree to any distribution of the profits]. It does not cost you anything to play the lotteries, but you are not allowed to refrain from playing at all. [*addition for group treatment only:* In case you cannot agree on one alternative, please leave the question unanswered. You will then receive the pre-specified standard payoff for this question.] In case you have difficulties to understand the structure of any of the lotteries, please do not hesitate to ask the supervisor of the experiment.

Example:

Lottery A pays 10,000 MU [30,000 MU] with a probability of 60 percent and 5,000 MU [15,000] with a probability of 40 percent.



**Decision 1: Please tick the box of the lottery that you [your team] prefer to play ( $X$  or  $X'$ )**  
*[addition for group treatment only: If you cannot find any mutual consent in the team, please leave out the question. You will then receive 0 MU if this lottery is selected to be played.]*

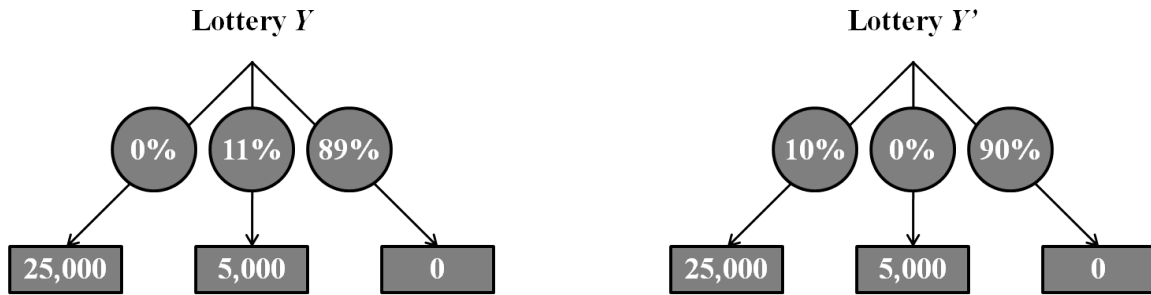
- ☐ **Lottery  $X$**  pays 5,000 MU [15,000 MU] with a probability of 100 percent (certainty).
- ☐ **Lottery  $X'$**  pays 25,000 MU [75,000 MU] with a probability of 10 percent, 5,000 MU [15,000 MU] with a probability of 89 percent and 0 MU with a probability of 1 percent.



*[payoffs in the pictures were changed according to the instructions above in the group treatment]*

**Decision 2: Please tick the box of the lottery that you [your team] prefer to play ( $Y$  or  $Y'$ )**  
*[addition for group treatment only: If you cannot find any mutual consent in the team, please leave out the question. You will then receive 0 MU if this lottery is selected to be played.]*

- ☐ **Lottery  $Y$**  pays 5,000 MU [15,000 MU] with a probability of 11 percent and 0 MU with a probability of 89 percent.
- ☐ **Lottery  $Y'$**  pays 25,000 MU [75,000 MU] with a probability of 10 percent and 0 MU with a probability of 90 percent.



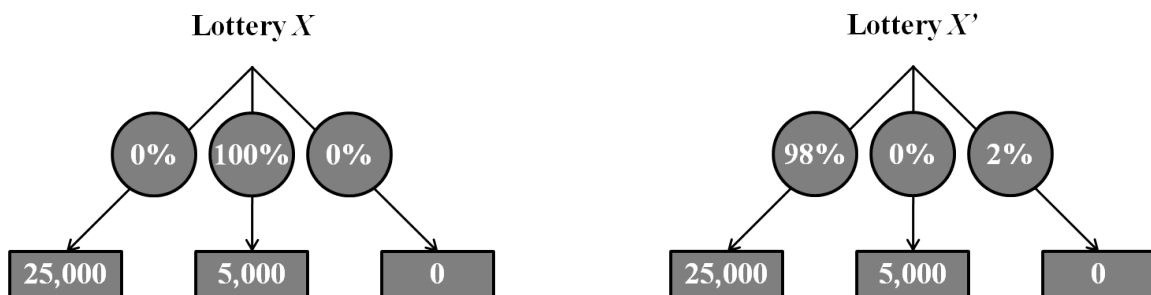
[payoffs in the pictures were changed according to the instructions above in the group treatment]

Allais Paradox (common ratio), between lotteries 11A and 12 (see Appendix 2.C)

(the introduction and the example from the common consequences version as stated above were also shown before this version)

**Decision 3: Please tick the box of the lottery that you [your team] prefer to play (*X* or *X'*)**  
 [addition for group treatment only: If you cannot find any mutual consent in the team, please leave out the question. You will then receive 0 MU if this lottery is selected to be played.]

- ☐ **Lottery *X*** pays 5,000 MU [15,000 MU] with a probability of 100 percent (certainty).
- ☐ **Lottery *X'*** pays 25,000 MU [75,000 MU] with a probability of 98 percent and 0 MU with a probability of 2 percent.



[payoffs in the pictures were changed according to the instructions above in the group treatment]



**Decision 4: Please tick the box of the lottery that you [your team] prefer to play (*Y* or *Y'*)**  
*[addition for group treatment only: If you cannot find any mutual consent in the team, please leave out the question. You will then receive 0 MU if this lottery is selected to be played.]*

- ☐ **Lottery *Y*** pays 5,000 MU [15,000 MU] with a probability of 1 percent and 0 MU with a probability of 99 percent.
- ☐ **Lottery *Y'*** pays 25,000 MU [75,000 MU] with a probability of 0.98 percent and 0 MU with a probability of 99.02 percent.



*[payoffs in the pictures were changed according to the instructions above in the group treatment]*

#### Ellsberg Paradox, between lotteries 15A and 16 (see Appendix 2.C)

In the following, we again show you two pairs of lotteries. To help you to understand the lotteries, they are described as urns that contain balls in different colors. The colors represent different potential payoffs (in MU). The frequencies with which the colors occur illustrate the probabilities to win the respective payoffs.

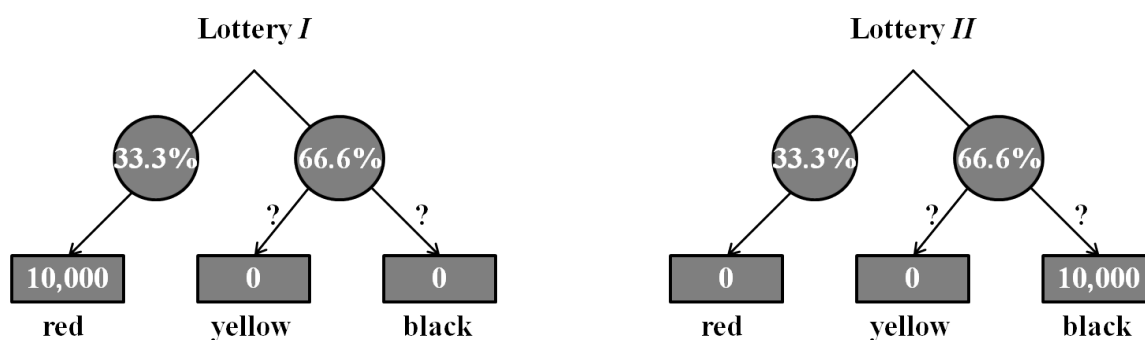
You are asked to decide for the lottery that you would like to play better. *[addition for group treatment only: You shall make an unanimous decision as a team, i.e. each team member has to agree to the choice].* Lotteries are described verbally and graphically. The probabilities of the outcomes are stated in percentages and the value of the outcomes is stated in “monetary units” MU *[addition for group treatment only: and are to be understood as total payoffs for your team. You may agree to any distribution of the profits].* It does not cost you anything to play the lotteries, but you are not allowed to refrain from playing at all. *[addition for group treatment only: In case you cannot agree on one alternative, please leave the question unanswered. You will then receive the pre-specified standard payoff for this question.]* In case

you have difficulties to understand the structure of any of the lotteries, please do not hesitate to ask the supervisor of the experiment.

Imagine an urn containing 90 balls of three different colors, red, yellow, and black. The number of red balls is exactly 30. All remaining balls are either yellow or black, with an unknown distribution. The outcome of the two lotteries depends on the color of the drawn ball.

**Decision 5: Please tick the box of the lottery that you [your team] prefer to play (*I* or *II*)**  
*[addition for group treatment only: If you cannot find any mutual consent in the team, please leave out the question. You will then receive 0 MU if this lottery is selected to be played.]*

- ☐ In **Lottery I**, a red ball pays 10,000 MU [30,000 MU], while yellow and black balls pay 0 MU.
- ☐ In **Lottery II**, a black ball pays 10,000 MU [30,000 MU], while yellow and red balls pay 0 MU.

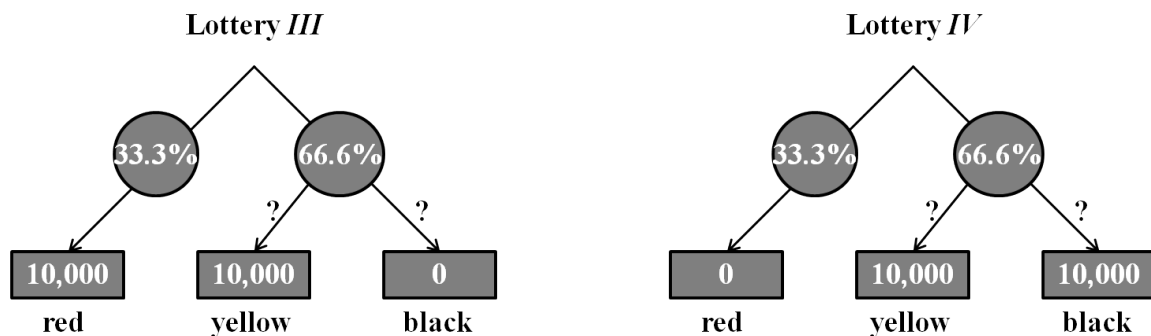


[payoffs in the pictures were changed according to the instructions above in the group treatment]

Imagine the very same urn as in the previous decision 5, containing exactly the same constellation of 90 colored balls, i.e. 30 red balls and – even though still not known – the same number of yellow balls and the same number of black balls as before. Again, the outcome of the two lotteries depends on the color of the drawn ball.

**Decision 6: Please tick the box of the lottery that you [your team] prefer to play (*III* or *IV*) [addition for group treatment only: If you cannot find any mutual consent in the team, please leave out the question. You will then receive 0 MU if this lottery is selected to be played.]**

- ☐ In **Lottery *III***, red and yellow balls pay 10,000 MU [30,000 MU], while black balls pay 0 MU.
- ☐ In **Lottery *IV***, yellow and black balls pay 10,000 MU [30,000 MU], while red balls pay 0 MU.



[payoffs in the pictures were changed according to the instructions above in the group treatment]

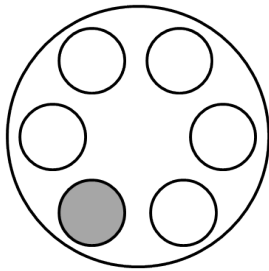
#### Zeckhauser Paradox (Russian Roulette), between lotteries 20A and 21 (see Appendix 2.C)

The following task is a pure thought experiment. Imagine you are forced to play “Russian Roulette”. It is played with a revolver with six chambers. By spinning the cylinder, either a loaded chamber or an empty chamber is randomly selected. If the chamber is empty, you survive [your whole team survives]. If the chamber is loaded, the game ends fatally for you [for your whole team].

Imagine that your total wealth has a value of 10,000 MU [Imagine the total wealth of all team member has a value of 30,000 MU and is owned by all members in equal shares, i.e. every team member is equally rich]. The revolver is loaded with one bullet. It is offered to you that you can buy the removal of the bullet. How much is it worth to you (in MU) to increase your probability of surviving the game to 100 percent by removing the bullet? [addition for group treatment only: Each team member has to contribute equally to the payment, i.e. you have to agree on a total amount that is acceptable to every member. Please report the total amount for your team, not each team member’s share!]

**Decision 7: Please indicate your maximal willingness to pay in MU to have the bullet removed [addition for group treatment only: If you cannot find any mutual consent in the team, please leave out the question].**

Please report an integer value (e.g. 537 or 8600, not 2500.3)



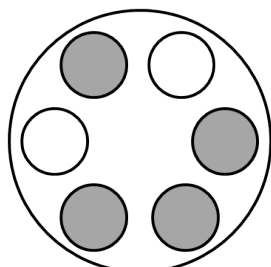
I [We] would maximally pay \_\_\_\_\_ MU to have the bullet removed.

The following task is again a pure thought experiment. Imagine you are forced to play “Russian Roulette”. It is played with a revolver with six chambers. By spinning the cylinder, either a loaded chamber or an empty chamber is randomly selected. If the chamber is empty, you survive [your whole team survives]. If the chamber is loaded, the game ends fatally for you [for your whole team].

Imagine that your total wealth has a value of 10,000 MU [Imagine the total wealth of all team member has a value of 30,000 MU and is owned by all members in equal shares, i.e. every team member is equally rich]. The revolver is loaded with four bullets. It is offered to you that you can buy the removal of one bullet. How much is it worth to you (in MU) to increase your probability of surviving the game to 50 percent by removing one bullet? [addition for group treatment only: Each team member has to contribute equally to the payment, i.e. you have to agree on a total amount that is acceptable to every member. Please report the total amount for your team, not each team member’s share!]

**Decision 8: Please indicate your maximal willingness to pay in MU to have one bullet removed [addition for group treatment only: If you cannot find any mutual consent in the team, please leave out the question].**

Please report an integer value (e.g. 537 or 8600, not 2500.3)



I [We] would maximally pay \_\_\_\_\_ MU to have one bullet removed.

Zeckhauser Paradox (Dice Game), after lottery 24A (see Appendix 2.C)

Before the game, you are endowed with an amount of 10,000 MU [30,000 for your whole team]. The throw of a die determines whether you may keep the endowment. If the die shows 1 pip, you lose the total amount of 10,000 MU. If the die shows 2, 3, 4, 5 or 6 pips, you may keep the total amount of 10,000 MU. Which amount  $X$  are you maximally willing to pay in order to have these rules changed in a way that you can keep the remaining amount (10,000 MU –  $X$ ) when the die shows 1, 2, 3, 4, 5 or 6 pips – thus with a probability of 100 percent? [addition for group treatment only: Each team member has to contribute equally to the payment, i.e. you have to agree on a total amount that is acceptable to every member. Please report the total amount for your team, not each team member's share!]

**Decision 9: Please indicate your maximal willingness to pay [addition for group treatment only: If you cannot find any mutual consent in the team, please leave out the question. You will then lose the total amount of 30,000 MU if this lottery is selected to be played].**

Please report an integer value (e.g. 537 or 8600, not 2500.3)

I [We] would maximally pay \_\_\_\_\_ MU.

Before the game, you are endowed with an amount of 10,000 MU [30,000 for your whole team]. The throw of a die determines whether you may keep the endowment. If the die shows

1, 2, 3 or 4 pips, you lose the total amount of 10,000 MU. If the die shows 5 or 6 pips, you may keep the total amount of 10,000 MU. Which amount X are you maximally willing to pay in order to have these rules changed in a way that you can keep the remaining amount (10,000 MU – X) when the die shows 4, 5 or 6 pips – thus with a probability of 50 percent? [*addition for group treatment only*: Each team member has to contribute equally to the payment, i.e. you have to agree on a total amount that is acceptable to every member. Please report the total amount for your team, not each team member's share!]

**Decision 10: Please indicate your maximal willingness to pay [*addition for group treatment only*: If you cannot find any mutual consent in the team, please leave out the question. You will then lose the total amount of 30,000 MU if this lottery is selected to be played].**

Please report an integer value (e.g. 537 or 8600, not 2500.3)

I [We] would maximally pay \_\_\_\_\_ MU.

**Appendix 2.E / Questionnaire****Name:** \_\_\_\_\_**Group No.:** \_\_\_\_\_

<b>1</b>	<p>Please evaluate the behavior of the other test persons in your group during the discussion on the scales below. Please provide evaluations for each group member, and on all scales. Each scale describes an attribute, with a value of -3 indicating that the person shows a very low degree of that behavior, 0 indicating an average degree, and +3 corresponding to a very high degree. Values in between stand for degrees slightly below and above average (-1 and 1) or considerably below and above average (-2 and 2). Your judgment will be kept anonymously, i.e. the other group members will not be able to find out how you have judged them. Neither your own name nor the names of any of your group members will be published.</p>																		
<b>1a)</b>	<p>Name of group member 1: _____</p>																		
<p>The analytical competence of group member 1 was...</p> <table> <tr> <td></td> <td>-3</td> <td>-2</td> <td>-1</td> <td>0</td> <td>1</td> <td>2</td> <td>3</td> <td></td> </tr> <tr> <td>very weak</td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td>very strong</td> </tr> </table>			-3	-2	-1	0	1	2	3		very weak	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	very strong
	-3	-2	-1	0	1	2	3												
very weak	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	very strong											
<p>The goal-orientation of group member 1 was...</p> <table> <tr> <td></td> <td>-3</td> <td>-2</td> <td>-1</td> <td>0</td> <td>1</td> <td>2</td> <td>3</td> <td></td> </tr> <tr> <td>very weak</td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td>very strong</td> </tr> </table>			-3	-2	-1	0	1	2	3		very weak	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	very strong
	-3	-2	-1	0	1	2	3												
very weak	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	very strong											
<p>The persuasiveness of group member 1 was...</p> <table> <tr> <td></td> <td>-3</td> <td>-2</td> <td>-1</td> <td>0</td> <td>1</td> <td>2</td> <td>3</td> <td></td> </tr> <tr> <td>very weak</td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td>very strong</td> </tr> </table>			-3	-2	-1	0	1	2	3		very weak	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	very strong
	-3	-2	-1	0	1	2	3												
very weak	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	very strong											
<p>The leadership role of group member 1 was...</p> <table> <tr> <td></td> <td>-3</td> <td>-2</td> <td>-1</td> <td>0</td> <td>1</td> <td>2</td> <td>3</td> <td></td> </tr> <tr> <td>very weak</td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td>very strong</td> </tr> </table>			-3	-2	-1	0	1	2	3		very weak	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	very strong
	-3	-2	-1	0	1	2	3												
very weak	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	very strong											
<b>1b)</b>	<p>Name of group member 2: _____</p>																		
<p>The analytical competence of group member 2 was...</p> <table> <tr> <td></td> <td>-3</td> <td>-2</td> <td>-1</td> <td>0</td> <td>1</td> <td>2</td> <td>3</td> <td></td> </tr> <tr> <td>very weak</td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td>very strong</td> </tr> </table>			-3	-2	-1	0	1	2	3		very weak	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	very strong
	-3	-2	-1	0	1	2	3												
very weak	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	very strong											
<p>The goal-orientation of group member 2 was...</p> <table> <tr> <td></td> <td>-3</td> <td>-2</td> <td>-1</td> <td>0</td> <td>1</td> <td>2</td> <td>3</td> <td></td> </tr> <tr> <td>very weak</td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td>very strong</td> </tr> </table>			-3	-2	-1	0	1	2	3		very weak	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	very strong
	-3	-2	-1	0	1	2	3												
very weak	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	very strong											
<p>The persuasiveness of group member 2 was...</p> <table> <tr> <td></td> <td>-3</td> <td>-2</td> <td>-1</td> <td>0</td> <td>1</td> <td>2</td> <td>3</td> <td></td> </tr> <tr> <td>very weak</td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td>very strong</td> </tr> </table>			-3	-2	-1	0	1	2	3		very weak	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	very strong
	-3	-2	-1	0	1	2	3												
very weak	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	very strong											

The leadership role of group member 2 was...									
		-3	-2	-1	0	1	2	3	
		very weak	○	○	○	○	○	○	very strong
2	Are you familiar with the Allais-paradox (e.g. from lectures or textbooks)?	<input type="checkbox"/> yes	<input type="checkbox"/> no						
3	Are you familiar with the Ellsberg-paradox (e.g. from lectures or textbooks)?	<input type="checkbox"/> yes	<input type="checkbox"/> no						
4	Are you familiar with the Zeckhauser paradox (e.g. from lectures or textbooks)?	<input type="checkbox"/> yes	<input type="checkbox"/> no						
5	Please indicate whether you agree to the following statement regarding a decision you have to make: When comparing the available lotteries, I rank them in a way that it is possible for me to say which one of two specific lotteries I prefer or if they are equivalently valuable for me.	<input type="checkbox"/> I agree	<input type="checkbox"/> I am not sure	<input type="checkbox"/> I do not agree					
6	Assume that you prefer lottery A over lottery B. In turn, you prefer lottery B over lottery C. Please indicate whether you agree to the following statement regarding a decision you have to make: I prefer lottery A over lottery C.	<input type="checkbox"/> I agree	<input type="checkbox"/> I am not sure	<input type="checkbox"/> I do not agree					
7	Assume that you prefer lottery A over lottery B. Please indicate whether you agree to the following statement regarding a decision you have to make: As long as the odds in both lotteries change only marginally, I stick to my preferences and prefer lottery A over lottery B.	<input type="checkbox"/> I agree	<input type="checkbox"/> I am not sure	<input type="checkbox"/> I do not agree					
8	Assume that you prefer lottery A over lottery B. Please indicate whether you agree to the following statement regarding a decision you have to make: Adding an identical chance of winning (i.e. with the same amount and the same probability) to both lotteries A and B does not change preference order and I still prefer lottery A over lottery B.	<input type="checkbox"/> I agree	<input type="checkbox"/> I am not sure	<input type="checkbox"/> I do not agree					
9	Assume that you prefer lottery A over lottery B, because lottery A yields a better result in the state of nature called "normal". In the case that any other state of nature materializes, you only know that both lotteries yield identical results, i.e. A is always as good as B. Please indicate whether you agree to the following statement regarding a decision you have to make: additional information on the other states of nature, for example their probability of occurrence or the (identical!) amounts of gains and losses in these states, does not change my preference order. I still prefer lottery A over lottery B.	<input type="checkbox"/> I agree	<input type="checkbox"/> I am not sure	<input type="checkbox"/> I do not agree					



<b>10</b>	What is your grade of your Abitur (please indicate by the first decimal, e.g. "1.8")?	_____	<input type="checkbox"/> other graduation
<b>11</b>	What is your age?	_____ years	
<b>22</b>	What is your gender?	<input type="checkbox"/> female	<input type="checkbox"/> male
<b>13</b>	What is your nationality?	_____	

# 3 Risk Disclosure in Key Investor Information Documents, Risk Perception, and Investment Decisions<sup>38</sup>

*“[...] people choose, in effect, between descriptions of options rather than between the options themselves.” Amos Tversky<sup>39</sup>*

## 3.1 Introduction

Since July 1, 2011, it is mandatory for UCITS (Undertakings for Collective Investments in Transferable Securities)<sup>40</sup> in the European Union (EU) to publish a “Key Investor Information Document” (“KID” or “KIID”) that contains standardized information about the investment vehicle, namely

- i. an overview on the goals and the investment policy,
- ii. the risk and reward profile, consisting of a numerical scale and a verbal description,
- iii. costs that occur at the purchase and divestiture, as well as during the holding period,
- iv. the past performance and its corresponding benchmark (if any), as well as
- v. a section on other practical information containing references to applicable law, further information material, the regulatory authority in charge and the managing institution.

While KIDs are mandatory for UCITS only, most other financial products, like stocks, bonds, financial futures, options, certificates, etc. are equipped with product information sheets (in German law these are called Produktinformationsbogen, in short “PIB”) that essentially serve the same purpose, namely to enhance transparency and comparability of these products. One crucial difference between KIDs and PIBs, however, is that KIDs contain a standardized risk scale, the so-called “Synthetic Risk and Reward Indicator” (“SRRI”), which maps a fund’s

---

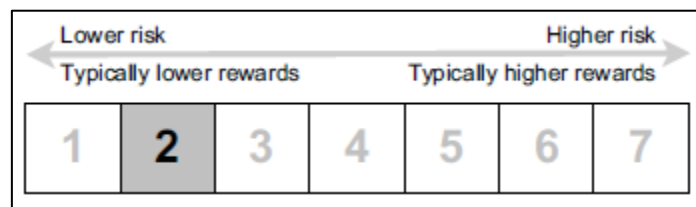
<sup>38</sup> The original essay, entitled “Risk Disclosure in Key Investor Information Documents, Risk Perception, and Investment Decisions”, is joint research work of Lutz Johanning and me. See Johanning and Troßbach (2014), unpublished manuscript.

<sup>39</sup> See Tversky (1996), page 7. The quote is taken from a discussion of description invariance, a core assumption of the theory of rational choice, that is frequently found to be violated in actual behavior, e.g. through framing effects.

<sup>40</sup> The expression roots back to a European Community directive from 1985 (85/611/EEC) that pioneered the common regulation of investment funds in all member states.

risk and reward profile into a single number. PIBs only provide verbal descriptions of the risk factors that are relevant for the financial product in question. Furthermore, there are exemplary scenario simulations that show contingent payoffs.

Figure 3.1: Synthetic Risk and Reward Indicator used in KIDs



The methodological standards to calculate a fund’s specific SRRI value were developed by the Committee of European Securities Regulators (CESR) for the European Commission before the newly-installed European Securities and Markets Authority (ESMA) took over the responsibility for UCITS regulation in 2011. Detailed instructions can be found in “CESR’s guidelines on the methodology for the calculation of the synthetic risk and reward indicator in the Key Investor Information Document” (CESR 2010). In short, the volatility of a fund’s weekly returns over the past five years is classified according to threshold values that determine the SRRI value of the fund under consideration. The threshold values are, starting from the maximum volatility of a fund with an SRRI value of 1 and ending with the minimum volatility of a fund having a value of 7: 0.5%, 2%, 5%, 10%, 15%, and 25%. Furthermore, the document specifies how migrations of SRRI values over time shall be addressed and how volatility shall be computed in cases where the standard methodology is not feasible, i.e. due to lack of data or the specific nature of a fund. The funds that we use in our experiment are all classified according to the standard procedure.

In this chapter we analyze how the variation of the descriptions of the risk and reward profiles influences risk perception and investment decisions. The SRRI indicates a fund’s historical volatility over the last five years, while the verbal explanation lists other risk factors that are not sufficiently reflected by this scale if they are relevant for the fund in question. These are credit risk, liquidity risk, default risk of third parties, operational risk, and risks associated with the use of derivatives. Furthermore, we present evaluations of the usability of KIDs and derive proposals for further improvement. We will also discuss a potential amendment of PIBs through the addition of a risk scale.

By enacting the Investor Protection and Improvement of the Functioning of Capital Markets Act (Gesetz zur Stärkung des Anlegerschutzes und Verbesserung der Funktionsfähigkeit des Kapitalmarktes (AnSFuG)), the German legislative constituted the legal basis for KIDs by implementing EU regulation No. 583/2010 in federal law. The AnSFuG introduces a number of changes and extensions to existing German law codes, with the Securities Trading Act (Wertpapierhandelsgesetz (WpHG)) and the Investment Act (Investmentgesetz (InvG)) containing the detailed objectives and requirements regarding KIDs. They are intended to enable private investors to easily understand and compare different UCITS in order to arrive at a sound investment decision. This shall be achieved by standardization, briefness, comprehensibility and clearness. The law applies to all forms of UCITS, which are primarily traditional investment funds and exchange-traded funds (ETFs), and makes further specifications for other fund structures like real estate investment trusts (REITs), funds of funds, master-feeder fund structures, funds with several tranches or share classes, funds with capital protection etc. In this study, we focus on plain investment funds. Financial advisors are obliged to hand out KIDs of funds that a client considers for an investment before the actual decision is made. They are also publicly available on the websites of the issuing investment companies and through numerous other information sources. In all other European Union member states, essentially identical legislations are installed. Therefore our results and conclusions can easily be generalized to the EU-wide investment fund market. However, in this article, we refer to German law and the funds that we analyze are managed by a German investment company.

We conduct an experimental study in which subjects are asked to allocate a fictitious amount of EUR 10,000 to four different investment funds and a no-interest checking account. The checking account serves as non-investment option to control for the tendency to participate in capital markets at all. We vary the information set that is available to the subjects about the risk and reward profiles in four different treatments. In the base treatment A, we provide the original KIDs of the funds, containing both the scale and verbal descriptions. In treatment B, we take out the scale measure and provide the verbal risk description only, as it is done in PIBs. The third treatment C shows only the risk scale without any further verbal content. Finally, in the fourth treatment D, the whole section on the risk and return profile is taken out of the KIDs. Here, subjects can still observe differences in the riskiness of funds by comparing the past performance that is represented by bar charts depicting their annual performance over the last ten years, but distinction between the funds is much less clear-cut. It

has been extensively discussed in the literature that mutual fund investors mainly pay attention to a fund's past performance when making their investment decisions (see for example Brown et al. (1996) or Busse (2001)). Therefore, this treatment may be considered to capture typical investment behavior that would be observable without the introduction of KIDs, especially the standardized description of risk and reward profiles. We conducted an online experiment with a hypothetical investment decision rather than a field study, because it would be a breach of the effective regulation to modify the KIDs in the way we did when using them for real investment decisions.

We find that risk perception becomes significantly biased when the scale is removed and subjects only have the verbal descriptions at hand. The least and the most risky investment option are identified correctly less frequently. The verbal description of risk factors is difficult to interpret if it is considered separately and subjects can easily get confused by this. When we remove the whole section on the risk and reward profiles of the funds, risk perception is also significantly biased compared to the case with full KIDs, but notably not as much as in the case where a verbal description is provided. We also observe various differences in actual investment decisions that are driven by the experimental variations. Providing information on the risk and reward properties of funds increases the tendency to invest especially with persons who have little financial knowledge and experience. Furthermore, if investors are better able to identify the most and the least risky fund, the invested amounts into these funds change significantly, and this also changes the overall portfolio risk.

## **3.2 Experimental Design**

### **3.2.1 Reference Study from the US market**

The study which is closest in essence to ours is that of Heinberg et al. (2010). Similar to our approach, the authors vary the ways in which the riskiness of different investment funds is presented. We start this section by discussing the general differences between the tools applied in the US and EU to disclose fund risk and reward profiles, the experimental design and the results by Heinberg and co-authors, and we derive some important implications for our study.

Heinberg et al. (2010) use the US Labor Department's "Model Comparative Chart" (MCC) as their reference model, which differs in two important attributes from KIDs. First, the MCC in its official format does not have a separate section or item devoted to the quantification or

description of risks. Instead, it only provides information on the funds' and their benchmarks' average annual returns for the last year, over the past five and ten years, and the total time period since inception. In KIDs, the annual fund and benchmark returns for each of the past ten years is reported graphically in a bar chart. Heinberg et al. (2010) modify the standard MCC in their experiment by adding two different items to describe the risk and reward profile of each investment option. Second, while KIDs are fund-specific and designed to compare each fund with the whole universe of other funds, the MCC contains a pre-selection of different funds that are available to choose from for any given retirement plan in the US. Usually, the options are listed in the order of their typical riskiness (i.e. the long-term average volatility by asset class). The authors identify significant order effects in the form that the first and the last option in the list receive the largest allocations. Order effects cannot distort investment decisions systematically when fund-specific KIDs are used. In our experiment, we randomize both, the order in which investment options are presented, and in which allocations are entered by the subjects.

In their study, Heinberg et al. (2010) introduce a summary rating of the risk and reward properties, which is in principle very similar to the SRRI, to the MCC. It is represented by a five-point scale where the number of shaded circles indicates the rating, for example one out of five for the least risky option and five out of five for the riskiest one. The rating is obtained by a ranking of the historical volatility of the funds' returns. See Figure 3.2 for an exemplary description of two funds with the highest and second-highest rating.

Figure 3.2: MCC with Additional Rating

Name/ Type of Option	Risk and Return Rating*  ●●●●● 1 2 3 4 5 Lower Higher	Average Annual Total Return as of 06/30/2010				Benchmark/Index as of 06/30/2010		
		1yr.	5yr.	10yr.	Since Inception	1yr.	5yr.	10yr.
Stock Funds								
Small Cap Stock Index Fund A	5 ●●●●●	22.7%	2.2%	3.7%	5.0% 05/21/98	22.9%	2.2%	3.6% Spliced Small Cap Stock Index
Stock Market Index Fund B	4 ●●●●○	14.3%	-0.9%	-1.7%	10.1% 08/31/76	14.4%	-0.8% S&P 500 Index	-1.6%



Source: Heinberg et al. (2010), Appendix 2, p.17

Subjects are not informed precisely about the calculation method of the rating, instead they are instructed that the “[r]isk and return ranking is determined by a reputable third party based on the variability of the historical annual returns of each option. On a scale from 1-5, lower ratings indicate less change in the fund’s performance over time (i.e. lower risk and reward)” (Heinberg et al (2010), p. 17).

Overall, this variation of the MCC does not have a significant impact on the allocations made by the participants of Heinberg et al.’s (2010) experiment (see the Tables 5 and 8 in their paper, p. 29 and 32): only the allocations to a money market fund with the rating one out of five were significantly higher when the risk and return rating was introduced. However, the authors also report that the introduction of the rating enables people to better identify the riskiness of the available options (see their Table 9 on p. 33).

The second variation of the MCC in the experiment of Heinberg et al. (2010) is to include a graphical display of past annual returns to each investment option. See Figure 3.3 below for an example.

Figure 3.3: Graphical Display of Annual Returns in MCC

Name/ Type of Option	Graph: 2000-2009 Year-End Total Returns*	Average Annual Total Return as of 06/30/2010				Benchmark/Index as of 06/30/2010		
		1yr.	5yr.	10yr.	Since Inception	1yr.	5yr.	10yr.
Stock Funds								
Small Cap Stock Index Fund A		22.7%	2.2%	3.7%	5.0% 05/21/98	22.9%	2.2%	3.6%
Stock Market Index Fund B		14.3%	-0.9%	-1.7%	10.1% 08/31/76	14.4%	-0.8%	-1.6%
						Spliced Small Cap Stock Index	S&P 500 Index	

Source: Heinberg et al. (2010), Appendix 2, p.17

Visual inspection suggests that it is hard to recognize the differences in the two options’ risk and reward profile. There is no axis labeling and generally the size of the graph is very small. Accordingly, the evidence for changes in asset allocation in this variation is mixed. Compared to the basic treatment without any additional information, two (four when controlling for order effects) of the investment options receive significantly different allocations, with the total riskiness of the portfolio being unchanged (see the Tables 5 and 8, p. 29 and 32 in

Heinberg et al. (2010)). Furthermore, the graphs are not helpful to recognize the volatility risk of the funds correctly (see Table 9 on p. 33 in Heinberg et al. (2010)).

### 3.2.2 Selection of Funds

For our experiment, we use KIDs of four investment funds that are managed by one of the largest German asset management companies. The subjects are informed that all information is taken from real world examples. However, we change everything that allows the participants to identify the funds unambiguously. First of all, we use fantasy names for the funds themselves (Fund BLUE, Fund GREEN, Fund RED, Fund WHITE), the investment management company (XY Fund Company S.A.), and the depository bank (ABC Bank Luxembourg S.A.). Second, we remove all company logos from the documents. Finally, we use nonexistent ISINs and cusip numbers.

As we focus on the effects of varying descriptions of the funds' risk and reward profiles, we selected the four funds in a way that they use the full range of the SRRI scale.<sup>41</sup> Fund BLUE invests in short term (average 12 months) corporate and government bonds and has an SRRI value of 1. Fund GREEN holds bonds with an investment period of 7 to 10 years, mainly issued by governments and other public institutions. Additionally, corporate bonds and emerging markets government bonds may be added to the portfolio. Fund GREEN has an SRRI value of 3. Fund RED invests in high yield corporate bonds and has an SRRI value of 5. Fund WHITE falls in the highest category with an SRRI value of 7. The fund mainly invests in European large-cap stocks, but also includes small- and mid-cap stocks in the portfolio. The stock selection is done by the fund managers along a number of qualitative and quantitative criteria. Furthermore, Fund WHITE is the only fund in our experiment that does not make use of derivatives. This has important implications for the verbal description of the fund's risk and return profile in the KID. Since Fund WHITE is the only investment option that is not exposed to default and counterparty risks of third parties other than the issuers of the non-derivative securities in the fund's portfolio, the verbal part only explains the volatility risk that is also depicted in the risk scale. Thus, in terms of the risk factors that are mentioned in Fund WHITE's KID, there is only a single source of risk (namely volatility), while the other three funds in our experiment, despite falling into lower risk categories, might seem riskier due to the mere presence of these additional risk factors. In this context, it is crucial for

---

<sup>41</sup> See Appendix 3.F, Appendix 3.G, Appendix 3.H, and Appendix 3.I for the full KIDs.



investors to understand especially the role of derivatives in the fund portfolios. Even though they introduce a new source of risk, they are usually used to counterbalance deviations from a fund's targeted risk and reward profile of the underlying investment strategy. They are used for risk management or term transformation purposes rather than for speculation. If investors have a different understanding of the role of derivatives and consider this an important factor in their risk and return tradeoff, the verbal risk description may have a distorting effect on allocations in funds that make use of derivatives. Our results clearly support this notion.

Of course it is impossible to find real world examples of funds with structurally different investment policies (and, hence, different risk and reward profiles) that are perfectly similar along all other dimensions described in the KID. For example, the fees that are charged from investors are typically higher in riskier funds. However, there are some important properties that all the funds in our experiment have in common.

- All funds invest in the Euro area only or predominantly (Fund GREEN may also invest minor amounts in emerging markets bonds) and are quoted in Euro in order to avoid distorting effects caused by the home bias (see e.g. French and Poterba (1991))
- All funds have a benchmark and an overall history of twenty years or more (inception dates range from 1988 to 1993; we ran the experiment in early 2013)
- The fact that all funds are managed by the same company ensures that the KIDs are similar in layout and color codes, which enables the test persons to recognize the uniformity of the documents at first glance

We are therefore confident that variations in our test persons' investment decisions should be mainly driven by their individual risk and reward preferences.

### 3.2.3 Experimental Stages

The experiment consists of the investment stage, the self-assessment stage and the KID assessment stage. The investment stage is varied over four different treatments; the other two stages are identical for all subjects. Each subject participates in one treatment only. In the beginning of the experiment, a short introduction and welcome screen informs the participants about the topic of the study, by whom and for what purpose it is conducted and that participating is voluntary. They are assured that all data is stored anonymously. Furthermore, participants are offered to take part in a lottery of fifteen vouchers worth 20 Euros each for the amazon online shop. At the end of the experiment subjects are instructed how they can

participate in the lottery.<sup>42</sup> The KID assessment stage is always the last stage, while the order of the investment stage and the self-assessment stage is randomized. The experiment closes with a short debriefing screen.<sup>43</sup>

### 3.2.4 Investment Stage<sup>44</sup>

In the investment stage the subjects are randomly assigned to one of the four experimental treatments. They are instructed to make their investment decision on the basis of the information that is provided in the experiment only and to imagine it would be a real decision using their own money. The task is to allocate a total amount of EUR 10,000 to four different investment funds and a no-interest checking account. The test persons are completely free to distribute the money over these five options as long as the full amount is allocated. Before the subjects enter their investment decision, they are instructed to carefully read the information documents of all four investment funds. These documents are the KIDs of the funds, with a different version of the “Risk and Reward Profile” section in each treatment. The participants are informed that the four documents contain standardized information about selected characteristics of the funds and that the fund management company is required by law to provide this information.

Treatment A uses the original form of the KIDs containing both, the standardized risk scale with the SRRI value of the fund and a verbal explanation of the other potential risk factors. There is also a supplementary verbal explanation on how to interpret the SRRI scale. We consider this an integral part of the scale.

In treatment B, we remove the risk scale (together with the supplementary explanation) from the KIDs and add the following to each fund’s verbal risk explanation: “The fund’s capital is invested in securities that are subject to certain price fluctuations. On the long-term historical average, larger fluctuations are accompanied by higher returns.” We mentioned before that Fund WHITE’s KID does not contain any verbal risk explanation in the original version due to the fact that there is no counterparty or default risk, because the fund portfolio only contains stocks. Therefore, if we had not added the general note on volatility risk, Fund WHITE’s “Risk and Reward Profile” section would have been empty while that of all other

---

<sup>42</sup> See Appendix to Chapter 3 for a translation of the full introduction and welcome screen.

<sup>43</sup> See Appendix 3.E for a translation of the debriefing screen.

<sup>44</sup> See Appendix 3.B for a translation of all questions in the investment stage.

funds would not. This is likely to cause suspiciousness with the subjects that take part in this treatment. Note that the additional sentences do not contain any quantitative information about the relative or absolute magnitudes of the respective price fluctuations. Therefore, we are confident that we do not materially influence the investment decision by this addition.

In treatment C, we remove the verbal explanation of potential other risk sources and provide the scales with the funds' SRRI values only (together with the supplementary information).

Finally, in treatment D, the section "Risk and Reward Profile" is left out completely.

All four KIDs in their respective form are opened as PDF files in separate browser windows. The order in which the links to the documents are displayed on the screen is randomized. The subjects are instructed to leave these windows open over the whole course of the experiment to have all information available at all times. After having studied the documents, subjects proceed to the actual investment decision. They are asked to enter a Euro value for each of the four funds and the checking account. Feasible values per entry range from 0 to 10,000. The sum of all entries has to be equal to 10,000. Empty entries are not allowed. These conditions are checked automatically and subjects receive an error message that asks them to complete all entries adhering to the instructions if at least one of them is violated. The order in which the five entry fields are displayed on the screen is also randomized. Therefore we do not have to control for any order effects in the allocations.

After the actual investment decision, the subjects are asked to rate a number of fund characteristics regarding their importance for the allocation of the EUR 10,000. Some of the items are directly related to elements of the KIDs, while others are more general. For each characteristic, there is five point Likert-scale ranging from 1 "very important" to 5 "unimportant" and an additional "don't know / no statement" option. The single characteristics are also listed in a randomized order. The last two entries are always blank and subjects can enter two more characteristics that they feel are missing in the catalogue. In detail, it consists of the following items:

- Investment policy of the funds
- Past financial performance of the funds
- Past value fluctuations of the funds
- Safety of the invested capital (asset preservation)
- Verbal descriptions of the risk and reward profiles of the funds (not for treatments C and D)

- Costs of the funds (one-time and recurring)
- Risk classes of the funds according to the Key Investor Information Documents (not for treatments B and D)
- Portfolio structure of the funds (distribution of the fund capital)

The final task in the investment stage is to rank the funds by their riskiness as it is perceived by each participant. The least risky fund shall be assigned a rank of 1, the second least risky fund is rank 2, then follows rank 3, and finally the riskiest fund receives a rank of 4. Again, the order in which the funds are displayed is randomized in this task.

### 3.2.5 Self-Assessment Stage<sup>45</sup>

In the self-assessment stage, the subjects are asked to provide some personal information about their experience with private investments and their risk attitude in the context of investing. First, the subjects are asked for their age, gender, and education level (measured by the highest degree obtained so far). Next, we want to know whether a participant is (or has been) involved with investments at work or whether she has studied or has been professionally trained in the field of investments. The test persons simply answer with “Yes” or “No”. After this, all participants are asked to evaluate their overall knowledge and experience in the context of investments, financial products, financial markets etc. A value of “1” indicates that a person regards herself completely financially illiterate. A value of “4” means that the person thinks she is as experienced as the population average. A value of “7” states that the person feels as financially literate as a professional investor. The last question in this stage asks for the personal risk attitude in the context of financial investments. Again, we provide a scale from 1 to 7, where a value of “1” means that a subject wants to be exposed to as little risk as possible and is therefore ready to forego return opportunities. A value of “4” is translated by “I am willing to take increased risks for higher return opportunities”. A value of “7” is in turn translated by “I am willing to take very high risks including the risk of total loss for very high return opportunities”. This final question is the reason why we randomize the order of the investment stage and the self-assessment stage: subjects might justify their investment decisions by categorizing themselves in a suitable risk category *ex post* and *vice versa*.

---

<sup>45</sup> See Appendix 3.C for a translation of all questions in the self-assessment stage.

### 3.2.6 KID Assessment Stage<sup>46</sup>

The KID assessment stage is always the last stage of the experiment. Here, the subjects are asked to evaluate the quality of the information that has been provided in the KIDs. On a scale from 1 to 5 (1 = “not agree at all”, 2 = “partly not agree”, 3 = “undecided”, 4 = “partly agree”, 5 = “fully agree”) the test persons shall indicate to what extent they agree with various statements concerning a) the suitability of the KIDs to compare the funds in different aspects, b) the importance of such comparisons in the investment decision, and c) the need for additional information or advice after reading the KIDs. The statements are always presented in the same order. In the next section, we present the results of our experiment.

## 3.3 Results

We start by presenting sample statistics of demographics, financial literacy, and risk tolerance that are collected in the self-assessment stage. Then we show how the participants evaluate the comparability of specific fund properties and how important it is for them to make these comparisons to arrive at an investment decisions. These items also serve as control variables in the analyses of risk perception and portfolio allocations. We conclude this section by summarizing the test persons’ evaluations of selected KID items and other factors that influence their decisions.

### 3.3.1 Sample Statistics

In total, 159 subjects completed the experiment, of which 41 (25.8 percent) are female and 118 (74.2 percent) are male. The average participant is 41.7 years old and the sample age ranges from 18 to 85 years. The subjects were assigned randomly and independently to the experimental treatments, which inevitably leads to unequal sample sizes. Treatment A was completed by 47 subjects, treatment B by only 36 subjects. Treatments C and D both had 38 participants. Except for just one, none of the variables that we measure in the different treatments is sufficiently normally distributed, as indicated by One-Sample Kolmogorov-Smirnov Tests. The divergences from normal distributions are all significant at least at a 95 percent confidence level. Only the allocation to Fund GREEN is sufficiently normally distributed in treatment C.<sup>47</sup> Therefore we apply non-parametric statistical tests throughout

---

<sup>46</sup> See Appendix 3.D for a translation of all questions in the KID assessment stage.

<sup>47</sup> Detailed test results are available from the authors upon request.

the whole chapter to analyze behavioral differences when making pairwise comparisons between treatments.

Table 3.1: Educational Level and Professional or Educational Backgrounds in Financial Investments

Highest Educational Attainment	n	thereof with a professional or educational background in financial investments
None	0	0
Certificate of Secondary Education	1	0 (0%)
General Certificate of Secondary Education	2	0 (0%)
Advanced Technical College Entrance Qualification	3	2 (67%)
General Qualification for University Entrance	33	18 (55%)
Polytechnic Degree	13	6 (46%)
University Degree	82	36 (44%)
Doctorate Degree	25	10 (40%)
Total	159	72 (45%)

Note: The translations may be unclear due to the specific structure of the educational system in Germany. For clarification, the German terms used in the questionnaire are (from top to bottom): Keiner / Hauptschulabschluss / Realschulabschluss / Fachhochschulreife / Allgemeine Hochschulreife / Fachhochschulabschluss / Hochschulabschluss / Dokortitel.

The educational level of the test persons is relatively high, as Table 3.1 illustrates. Furthermore, almost half of the participants have an educational or professional background in financial investments. These figures are probably driven by the fact that it is more likely for a person to take part in the experiment if she is interested in the topic. The fraction of professionally trained or educated people is fairly stable across educational levels.

When asked for a self-assessment of their financial literacy, the subjects categorize themselves as follows. The average classification is 4.245.

Table 3.2: Self-categorization in Financial Literacy

Self-Assessment in Financial Literacy	n	%
<b>1 (completely illiterate)</b>	13	8.2%
<b>2</b>	23	14.5%
<b>3</b>	11	6.9%
<b>4 (about the same as the population average)</b>	25	15.7%
<b>5</b>	46	28.9%
<b>6</b>	34	21.4%
<b>7 (like a professional investor)</b>	7	4.4%

The self-assessment of risk tolerance in financial investments is distributed as follows. The average risk tolerance is 3.478. It is important to point out here that the scale that we use for the self-assessment of the subjects' risk tolerance is technically not related to the SRRI scale, even though both range from 1 to 7 and higher values reflect higher riskiness. We cannot directly conclude from a person's risk tolerance which combination(s) of funds she should optimally choose in the investment stage. However, we should expect a positive relationship between risk tolerance and the average SRRI value of a subject's portfolio.

Table 3.3: Self-categorization in Risk Tolerance in Financial Investments

Self-Assessment in Risk Tolerance in Financial Investments	n	%
<b>1 I want to be exposed to as little risk as possible and I am ready to forego return opportunities</b>	15	9.4%
<b>2</b>	31	19.5%
<b>3</b>	33	20.8%
<b>4 I am willing to take increased risks for higher return opportunities</b>	37	23.3%
<b>5</b>	31	19.5%
<b>6</b>	10	6.3%
<b>7 I am willing to take very high risks including the risk of total loss for very high return opportunities</b>	2	1.3%

Given the gender imbalance and the high proportion of financially educated or professionally trained persons in our sample, it is worthwhile to calculate the means for financial literacy and risk tolerance assessments for the respective subsamples.

Table 3.4: Financial Literacy and Risk Tolerance by Professional / Educational Background and Gender

	full sample	professional / educational background in financial investments		female	male
		0	1		
<b>Self-Assessment in Financial Literacy</b>	4.245	3.483	5.167	2.902	4.712
<b>Self-Assessment in Risk Tolerance in Financial Investments</b>	3.478	3.092	3.944	2.732	3.737

Unsurprisingly, we observe a clearly higher self-reported financial literacy score for those subjects who have professional or educational experience in the field. They are also more prone to take on risk. Both differences are significantly different as indicated by Kolmogorov-Smirnov Tests (financial literacy:  $Z = 2.645$ ,  $p < 0.001$ ; risk tolerance:  $Z = 1.885$ ,  $p = 0.002$ ). This is interrelated with the gender imbalance in the sample, as 60 out of 72 subjects who have a professional or educational background in financial investments are male, whereas 58 of the 87 subjects who do not have a relevant background are female. A  $\chi^2$ -Test shows indeed that the two categories are not independent from each other (Pearson  $\chi^2 = 5.718$ ,  $p = 0.017$ ). Consequentially, financial literacy and risk tolerance are also significantly different across genders (financial literacy:  $Z = 2.516$ ,  $p < 0.001$ ; risk tolerance:  $Z = 1.927$ ,  $p = 0.001$ ). However, Table 3.5 shows that there are no structural differences between the four experimental treatments with regards to age, financial literacy, and risk tolerance. Therefore we can exclude that behavioral differences are driven by imbalances between the groups.



Table 3.5: Age, Financial Literacy, and Risk Tolerance by Treatment

	Average per Treatment			
	A	B	C	D
Age (U-Test)	39.766 <sub>a</sub>	44.500 <sub>a</sub>	41.658 <sub>a</sub>	41.395 <sub>a</sub>
Financial literacy (self-assessment) (KS-Test)	4.340 <sub>a</sub>	4.472 <sub>a</sub>	4.211 <sub>a</sub>	3.947 <sub>a</sub>
Risk tolerance (self-assessment) (KS-Test)	3.681 <sub>a</sub>	3.500 <sub>a</sub>	3.316 <sub>a</sub>	3.368 <sub>a</sub>

Note: Means which do *not* share the same subscript are significantly different from each other as indicated by either pairwise Mann Whitney U-Tests or Kolmogorov-Smirnov-Tests with a confidence level of 90 percent or higher. While the U-Test is the standard method to test for differences in non-parametrically distributed variables, the KS-Test is preferable whenever the included variables have a limited number of categories, which frequently leads to inconclusive rank orders in a U-Test. Detailed results of pairwise comparisons are available from the authors upon request.

### 3.3.2 Comparison of Fund Characteristics over Experimental Treatments

The intended purpose of a KID is primarily to enhance the comparability of different investment alternatives, especially investment funds. Therefore we ask the test persons to evaluate the usefulness of the provided information to compare a number of fund characteristics in the KID assessment stage. Even though this is the last stage of the experiment, we present the results before turning to the analyses of risk perception and investment decisions, because we use the data as control variables in the following subsections. The participants are asked to indicate their degree of consent to the statements presented in Table 3.6 on a scale from 1 (“not agree at all”) to 5 (“fully agree”).

Table 3.6: Usability of KIDs for Comparison of Fund Characteristics

	Average per Treatment			
	A	B	C	D
I was able to compare the objectives and investment policies of the funds without any problems (KS-Test)	3.128 <sub>a</sub>	3.389 <sub>a</sub>	3.421 <sub>a</sub>	3.000 <sub>a</sub>
It was particularly important for my investment decision to compare the objectives and investment policies of the funds (KS-Test)	3.511 <sub>a</sub>	4.167 <sub>b</sub>	3.500 <sub>a</sub>	3.737 <sub>a,b</sub>
I was able to compare the risk and reward profiles of the funds without any problems (KS-Test)	3.723 <sub>a</sub>	3.306 <sub>a</sub>	3.316 <sub>a</sub>	3.132 <sub>a</sub>
It was particularly important for my investment decision to compare the risk and reward profiles of the funds (KS-Test)	4.191 <sub>a</sub>	4.194 <sub>a</sub>	4.316 <sub>a</sub>	3.921 <sub>a</sub>
I was able to compare the past financial performances of the funds without any problems (KS-Test)	3.723 <sub>a</sub>	3.861 <sub>a</sub>	3.816 <sub>a</sub>	3.947 <sub>a</sub>
It was particularly important for my investment decision to compare the past financial performances of the funds (KS-Test)	3.745 <sub>a</sub>	3.750 <sub>a</sub>	3.868 <sub>a</sub>	3.579 <sub>a</sub>
I was able to compare the costs of the funds without any problems (KS-Test)	3.915 <sub>a</sub>	3.667 <sub>a</sub>	3.395 <sub>a</sub>	4.079 <sub>a</sub>
It was particularly important for my investment decision to compare the costs of the funds (KS-Test)	3.574 <sub>a</sub>	3.528 <sub>a</sub>	3.474 <sub>a</sub>	3.184 <sub>a</sub>
I was able to compare the funds in general without any problems (KS-Test)	3.298 <sub>a</sub>	3.472 <sub>a</sub>	3.289 <sub>a</sub>	3.000 <sub>a</sub>
It was particularly important for my investment decision to compare the funds in general (KS-Test)	4.106 <sub>a</sub>	4.306 <sub>a</sub>	4.237 <sub>a</sub>	4.079 <sub>a</sub>
Actually I would have needed to contact an investment advisor prior to my decision (KS-Test)	3.426 <sub>a</sub>	3.417 <sub>a</sub>	3.737 <sub>a</sub>	3.816 <sub>a</sub>
I would have liked to acquire further information about the available funds prior to my decision (KS-Test)	3.979 <sub>a</sub>	4.194 <sub>a</sub>	4.000 <sub>a</sub>	4.184 <sub>a</sub>

Note: Means which do not share the same subscript are significantly different from each other as indicated by pairwise Kolmogorov-Smirnov-Tests with a confidence level of 90 percent or higher. Detailed results of pairwise comparisons are available from the authors upon request. Additional descriptive statistics can be found in Appendix 3.K.

The comparison of the risk and reward profiles of the funds is the most important specific criterion for the participants in all treatments. Even though the comparability of this property is evaluated the highest in treatment A with the full KID, the differences between treatments are not significantly different from each other. Actually, the only item that is rated significantly differently over treatments is the importance to compare the objectives and investment policies. The subjects in those treatments where the SRRI is not available (treatments B and D) attach more importance to this than those in treatments A and C. The average score in treatment D is not significantly different from any other treatment, but the effect is clearly visible for treatment B. We suppose that the participants rely more on the information provided in the section on the investment policies of the funds to compensate for the missing information on the risk and reward profile. While this seems to work out sufficiently – the comparability of all listed fund characteristics is equal across all treatments – one could also argue contrarily that the provision of the standardized risk and reward scale draws the investor’s attention away from the – surely highly important and fundamental – information about the investment policy.

Overall, we see that across all treatments the comparability of the different fund properties is rated rather positive, but there is surely room for improvement. Also, the need for further information after reading the KIDs is quite high. Even though the test persons’ perception and evaluation of the provided information is not significantly changed by the experimental variations of the KIDs, we do observe differences in risk perception and investment decisions between treatments. We will discuss these in the next subsections.

### 3.3.3 Information Sets and Risk Perception

Before turning to the actual allocation to the funds and the checking account, we take a closer look at the perceived riskiness of the investment options and how it differs between the experimental treatments that provide different information sets about the funds’ risk and reward profiles in Table 3.7.

Smaller average ranks indicate less perceived riskiness of any given fund. The assigned ranks for the funds GREEN and RED are on average identical over treatments and overall correct (i.e. corresponding to the funds’ respective SRRI values), as indicated by the identical treatment medians and modes as well as pairwise Kolmogorov-Smirnov Tests for equal distributions. Fund BLUE (the least risky one according to the SRRI) is perceived significantly riskier when only the verbal risk description is available but the risk scale is not

(treatment B) compared to those treatments that include the SRRI value (treatments A and C). Fund WHITE (the riskiest one) on the other hand is perceived significantly less risky in treatment B compared to treatments A and C.

Table 3.7: Assigned Riskiness Rankings per Treatment

Assigned Rank Riskiness Fund BLUE	Treatment A (n=47)	Treatment B (n=36)	Treatment C (n=38)	Treatment D (n=38)
Mean	1.36	1.94	1.42	1.61
SE of mean	0.123	0.164	0.139	0.144
Median	1.00	2.00	1.00	1.00
Mode	1	1	1	1
Test for Differences	a	b	a	a,b
<b>Assigned Rank Riskiness Fund GREEN</b>				
Mean	2.04	2.14	2.08	2.16
SE of mean	0.068	0.179	0.095	0.149
Median	2.00	2.00	2.00	2.00
Mode	2	2	2	2
Test for Differences	a	a	a	a
<b>Assigned Rank Riskiness Fund RED</b>				
Mean	3.00	3.11	2.92	3.11
SE of mean	0.101	0.153	0.127	0.135
Median	3.00	3.00	3.00	3.00
Mode	3	3	3	3
Test for Differences	a	a	a	a
<b>Assigned Rank Riskiness Fund WHITE</b>				
Mean	3.60	2.81	3.58	3.13
SE of mean	0.120	0.186	0.134	0.169
Median	4.00	3.00	4.00	4.00
Mode	4	4	4	4
Test for Differences	a	b	a	a,b

Note: Assigned ranks in treatments *not* sharing the same letters in the rows “Test for Differences” are significantly different from each other as indicated by pairwise Kolmogorov-Smirnov tests with a confidence level of at least 90 percent. Detailed test results and frequency distributions of the assigned ranks per treatment and per fund are available from the authors upon request.

We conclude that the SRRI helps investors to differentiate the riskiness between funds on both ends of the scale, i.e. to correctly identify the least and most risky investment option, as they are less able to do so if the scale is taken away in treatment B. In medium-risk

investments we do not see a difference. The assigned ranks of funds BLUE and WHITE in treatment D (without risk-scale and without verbal description) are in between the ranks assigned in treatment B and in the two treatments that have the risk-scale available, but they are not significantly different from each of those. Thus, there is no significant overall difference in relative risk perception when taking away the whole section on risk (treatment D) compared to a full KID (treatment A), but leaving out the SRRI only leads to misconceptions at the tails of the risk spectrum. Thus, the verbal explanation by itself is rather confusing than helpful. For example, it is not straightforward to understand for an investor that the use of derivatives in those funds who invest in fixed income securities like Fund BLUE is due to hedging and term transformation purposes, i.e. to reduce risk. However, this becomes clearer when the SRRI value is taken into account. On the other hand, especially with the fluctuations of the graphically depicted annual returns of Fund RED being seemingly very high compared to Fund WHITE (at least it is difficult to rank them by their volatility at a first glance), the SRRI helps to gauge their relative magnitude, which is not quantified in the verbal description. Also keep in mind that Fund WHITE does not have any other risk factors mentioned in the verbal description aside from volatility. This might also lead to an underestimation of the fund's riskiness.

Aside from the observed differences in average rankings between treatments for the Funds BLUE and WHITE, there may also be individual drivers in place that influence a subject's risk perception, and in turn her ability to rank the funds correctly. Therefore, in the next step, we present some multivariate analyses in which we control for further individual characteristics. In order to measure a subject's ability to make consistent risk estimates, we calculate the Pearson correlation coefficient  $\rho_i^{risk}$  between the ranks that subject  $i$  assigns to each fund and the ranking of the four funds' SRRI values. If both rankings are identical,  $\rho_i^{risk}$  takes on a value of 1, reflecting the highest possible degree of consistency. On the other hand, a completely inverse ranking leaves the individual with a  $\rho_i^{risk}$  of -1. The average correlation coefficients for the four experimental treatments are presented in Table 3.8.

Table 3.8: Correlation between Assigned Fund Risk and SRRI Values

$\rho_i^{risk}$	Mean	Standard Error	90% Conf. Interval
Treatment A	0.766	0.077	0.638 – 0.894
Treatment B	0.356	0.088	0.210 – 0.501
Treatment C	0.732	0.086	0.580 – 0.873
Treatment D	0.553	0.086	0.411 – 0.695

Pairwise comparison of the observed correlations between treatments shows that risk estimates are significantly better in treatment A than in both, treatments B and D at confidence levels of more than 90 percent. Furthermore, subjects in treatment C make significantly better risk estimates than in treatment B at a confidence level of more than 99 percent. The observed differences between treatment D and treatments B and C are insignificant at the usual confidence levels of 90 percent or higher, but the resulting p-values are both between 10 and 15 percent.<sup>48</sup>

When using  $\rho_i^{risk}$  as the dependent variable in a multivariate analysis, we need to take into account that it is a limited dependent variable by nature of the correlation coefficient, which may lead to shortcomings in measuring the attribute that we are interested in, namely the ability to correctly estimate the fund risk. It may actually differ between two subjects who both have an observed  $\rho_i^{risk}$  of 1 or -1, respectively. To capture this, we apply a TOBIT model with a lower and an upper censored limit of the observed parameter  $\rho_i^{risk}$ :

$$\rho_i^{risk} = \begin{cases} y_i & \text{if } -1 < y_i < 1 \\ -1 & \text{if } y_i \leq -1 \\ 1 & \text{if } y_i \geq 1 \end{cases}$$

$y_i$  is the latent (i.e. the true, yet unobserved) variable that describes subject  $i$ 's ability to estimate fund risk consistently. The TOBIT model assumes that  $y_i$  can take on values outside the limits of the observed correlation coefficient. The regression coefficients in Table 3.9 show the influence of changes in the respective independent variable  $x_j$  on the latent variable

---

<sup>48</sup> Detailed results of the pairwise comparisons are available from the authors upon request.

$y_i$ , i.e.  $\beta_j = \frac{\partial E(y_i)}{\partial x_j}$ . Other marginal effects, for example the changes in the expected value of  $\rho_i^{risk}$ , are available from the authors upon request. The variables *Treatment B*, *C*, and *D*, as well as *Female* and *Professional* are dummy variables. Therefore, the corresponding  $\beta$ -coefficients reflect the expected change in  $y_i$  for a discrete change of the dummy from 0 to 1. All other independent variables are centered at their particular sample median. For example, the median age in the total sample is 33 years. A subject that is 35 years old is now assigned a value of 2 for her *age* variable, as she is two years older than the median person. This does not change the resulting estimates of the  $\beta$ -coefficients, but it has a nice implication for the interpretation of the constant: it reports the expected value of  $y_i$  with all other independent variables taking on a value of 0. With all treatment dummies being 0, we automatically consider treatment A, which in our model serves as the base category to which all other treatments are compared. However, instead of estimating the average parameter  $y_i$  of an individual from treatment A that is 0 years old and that has, for example, a risk tolerance of 0 (which is not even defined on the scale that we use), we rather apply the median values for all non-dummy control variables in our model. These are *Age*, *Financial Literacy*, *Risk Tolerance* (all obtained from the self-assessment stage), and two indices that measure a) the importance that subjects attach to the comparability of the funds and b) the need for further information and financial advice. These indices are calculated as average values of the variables *Importance of Comparability Investment Policy / Risk and Reward Profile / Performance / Cost / Overall*, and *Need for Consulting / More Information*, respectively. We use the indices instead of the constituent variables in order to keep the model simple and to avoid potential problems with multicollinearity. We leave out the evaluations of certain fund characteristics from the investment stage, as these are not fully available for all subjects, depending on the treatment and the use of the “no statement” option. Furthermore, we leave out the variables that describe the subjects’ abilities to compare certain fund characteristics when using the KIDs. The independent variable that is of interest in our analysis is a measure of a certain ability itself and it is not clear if the causal relationship between these characteristics is really unidirectional. Therefore, in order to avoid potential problems of endogeneity, and due to lack of proper instruments, we drop these items.

Furthermore, we incorporate interaction terms between the treatment dummy variables and selected control variables into our model to allow for the possibility that the influence of the

control variables differs across treatments.<sup>49</sup> Note that, compared to a standard regression model, this specification slightly changes the interpretation of the resulting main effect coefficients for those variables that are interacted with others in the model. They are now conditional main effects, indicating the influence of the respective independent variable on  $y_i$  conditional on *all other* main effect variables being 0.<sup>50</sup> Therefore, the differences in risk estimation between experimental treatments are not necessarily considered equal across, for example, all levels of importance that subjects assign to the comparability of funds. Instead, it is indicated by the sum of the conditional main effect and the interaction effect. Table 3.9 summarizes the results.

---

<sup>49</sup> We only report interaction terms if they are significant for at least one experimental treatment. All other control variables are not interacted in the model specification for the sake of tractability. It is sufficient to report the (regular) main effects only in these cases. Only the indices that measure the importance of fund comparability and the need for further information and advice turn out to exert treatment-specific influence on risk estimation quality.

<sup>50</sup> Remember that a value of 0 represents the sample median for all non-dummy variables. Thus, the conditional main effects of the treatment dummies report differences in risk estimation for subjects who have the same attributes as the respective medians.



Table 3.9: TOBIT Model (Interaction Effects) Risk Estimation

$y_i$	Interaction Effects			
	$\beta$	Robust SE	t-value	p-value
Constant	1.191***	0.181	6.58	0.000
Treatment B	-0.754***	0.175	-4.30	0.000
Treatment C	-0.010	0.202	-0.49	0.622
Treatment D	-0.344*	0.186	-1.85	0.067
Age (mc)	-0.005	0.004	-1.13	0.259
Female	-0.010	0.190	-0.05	0.958
Professional	0.038	0.191	0.20	0.844
Financial Literacy (mc)	0.046	0.061	0.74	0.459
Risk Tolerance (mc)	-0.030	0.062	-0.49	0.624
Importance of Comparability Index (mc)	-0.197	0.250	-0.79	0.430
Need for Information and Consulting Index (mc)	-0.015	0.125	-0.12	0.907
Treatment B * Importance of Comparability Index (mc)	0.280	0.299	0.94	0.350
Treatment C * Importance of Comparability Index (mc)	0.599*	0.353	1.70	0.092
Treatment D * Importance of Comparability Index (mc)	0.393	0.279	1.41	0.160
Treatment B * Need for Information and Consulting Index (mc)	-0.263*	0.151	-1.75	0.083
Treatment C * Need for Information and Consulting Index (mc)	-0.293*	0.167	-1.75	0.082
Treatment D * Need for Information and Consulting Index (mc)	-0.159	0.179	-0.89	0.376

Note: Number of observations = 159 (3 left-censored, 88 uncensored, 68 right-censored). Standard errors are robust against heteroskedasticity. \*\*\*, \*\*, and \* indicate statistical significance at the 99%, 95%, and 90% confidence levels. (mc) stands for median-centered.  $F(16, 143) = 4.06$ ,  $p < 0.001$ , Pseudo- $R^2 = 0.1108$ .

The most important result is that the coefficients for the *treatment B* and *D* dummies are significantly negative. *Treatment C* has a coefficient that is not significantly different from 0, meaning that the subjects in this treatment achieve the same degree of consistency in fund risk estimates than in treatment A, the base case. In pairwise comparisons we also test for differences between all three resulting regression coefficients for the treatment dummies through likelihood ratio tests. The coefficient for *treatment B* is significantly different from those for *treatment C* ( $F(1, 143) = 11.56$ ,  $p < 0.001$ ) and *treatment D* ( $F(1, 143) = 5.39$ ,  $p =$

0.0216), while there is no significant difference between the coefficients for *treatment C* and *D* ( $F(1, 143) = 1.61, p = 0.2072$ ). Thus, the experimental variation in treatment B, i.e. taking away the SRRI and leaving the test persons with verbal explanations only, causes a difference to the worse compared to all other treatments, also when controlling for other potentially influential factors. Note that the constant is positive and significant. This indicates that, on average, a subject in treatment A that has median attributes and priorities, is female, and does not have a relevant background from education or working experience, arrives at a highly consistent risk estimation, which is a very pleasant result. Taking away the verbal explanation and leaving investors with the SRRI only (treatment C) does not change the quality of the risk estimation, but with only the verbal part (treatment B) it worsens drastically. Even in treatment D the negative impact of taking away the entire section on the risk and reward profile ( $\beta = -0.318$ ) is clearly smaller than taking away the scale only and leaving the subjects with the verbal part ( $\beta = -0.712$ ). Thus, the verbal explanation confuses the test persons if it stands alone.

None of the conditional main effects for the control variables are significantly different from 0, but some of the interaction terms are. This means that the assigned importance of fund comparability and the need for information and advice actually do influence consistency of risk estimations, but only in specific experimental treatments. Only in treatment C (only the scale is provided), those subjects who consider the comparability of funds more important arrive at clearly more consistent estimates. If the scale is the only instrument at hand to compare the risk and reward profiles, it is fairly obvious that the subjects tend to rank the funds according to the scale if they find the comparison highly important.

On the other hand, it is only in treatments B and C that test persons who feel a higher need for further information and financial advice arrive at significantly less consistent risk estimates. Again, a verbal explanation alone is confusing investors, while a single number might leave the test persons with a feeling that too much abstraction has been done.

### 3.3.4 Portfolio Selection and Willingness to Invest

Having analyzed the differences in risk perception between the experimental treatments, we now turn to the actual investment decisions and the resulting portfolios. First, we show the average allocations to each of the four funds and the checking account per treatment, as well as the resulting average SRRI values of the portfolios, where the checking account is assigned an SRRI value of 0. Then, in an attempt to analyze under which experimental condition the

actual portfolio allocation fits best to the individual risk preferences, we present two different measures to show this. Table 3.10 summarizes the univariate results. Keep in mind that subjects in treatment B have a significantly biased perception of the funds' riskiness in a way that the risk of Fund BLUE is overestimated, while Fund WHITE is underestimated.

Table 3.10: Allocations per Treatment

	Average per Treatment			
	A	B	C	D
<b>Checking Account (U-Test)</b>	1372.340 <sub>a</sub>	1479.167 <sub>a,b</sub>	1342.105 <sub>a</sub>	2368.421 <sub>b</sub>
<b>Fund BLUE (U-Test)</b>	2491.489 <sub>a</sub>	1256.944 <sub>b</sub>	1592.105 <sub>a,b</sub>	2460.526 <sub>a</sub>
<b>Fund GREEN (U-Test)</b>	2814.894 <sub>a,b,c</sub>	2236.111 <sub>b,d</sub>	3447.368 <sub>c</sub>	1934.211 <sub>d</sub>
<b>Fund RED (U-Test)</b>	1974.468 <sub>a</sub>	1708.333 <sub>a</sub>	2013.158 <sub>a</sub>	1578.947 <sub>a</sub>
<b>Fund WHITE (U-Test)</b>	1346.809 <sub>a</sub>	3319.444 <sub>b</sub>	1605.263 <sub>a</sub>	1657.895 <sub>a</sub>
<b>Value weighted SRRI of Allocation (U-Test)</b>	3.024 <sub>a,b</sub>	3.974 <sub>c</sub>	3.324 <sub>a,c</sub>	2.776 <sub>b</sub>
<b>Weighted SRRI minus Risk Tolerance (U-Test)</b>	-0.657 <sub>a</sub>	0.474 <sub>b</sub>	0.008 <sub>a,b</sub>	-0.592 <sub>a</sub>
<b>Mean</b>				
<b>Weighted SRRI minus Risk Tolerance</b>	-0.5	0.2	-0.2	-0.4
<b>Median (Full sample: -0.3)</b>		**		*
	++			+++

Note: Means which do not share the same subscript are significantly different from each other as indicated by pairwise Mann Whitney U-Tests with a confidence level of 90 percent or higher. Detailed results of pairwise comparisons are available from the authors upon request. In the last row, \*\*\*, \*\*, and \* indicate significant deviations of the treatment medians from the full sample median -0.3 with a confidence level of 99, 95, and 90 percent, respectively, as indicated by Wilcoxon Signed Rank Tests. +++, ++, and + indicate significant differences of the treatment medians from 0 with a confidence level of 99, 95, and 90 percent, respectively, as indicated by Wilcoxon Signed Rank Tests.

Looking at the allocations to the interest-free checking account in the first row of Table 3.10, we see that subjects in treatment D invest significantly less in the four funds than in treatments A and C. On average, the amount allocated to the checking account in treatment B is also much lower than in treatment D, however, the difference is not statistically significant according to a Mann Whitney U-Test. There are no significant differences between the average amounts in treatments A, B, and C. An explanation for this behavior could be the reduction of uncertainty about fund risk which drives the overall tendency to invest. However, the multivariate analysis in Table 3.11 reveals that the drivers of overall participation are not pure treatment effects. We regress the allocation to the checking account on a number of

control variables including dummies for treatments B to D, i.e. treatment A serves as base category. The other controls are *age*, *gender*, *professional*, *financial literacy*, and *risk tolerance*, as well as the comparability evaluations of the different fund properties and the need for more information and consulting. The latter two are summed up as indices in analogy to the model presented in Table 3.9. In addition to the main effects, we also report significant interaction effects between the treatment dummies and the control variables.

None of the resulting coefficients for the conditional main treatment effects are significantly different from zero. Furthermore, pairwise comparisons of the resulting coefficients for the treatment dummy variables do not show any significant differences.<sup>51</sup> Instead, we see that the decision to leave money on the checking account over all four treatments is significantly driven by the comparability of the funds, with subjects who are better able to compare the available investment options investing more and leaving less money on the checking account. There are also two significant interaction terms: While the degree of financial literacy does not generally influence the tendency to invest (the conditional main effect is insignificant), in treatment D it does. The higher amounts on the checking account that we see in Table 3.10 for treatment D are driven by those subjects who are financially illiterate. With each additional score point in the self-assessment of financial literacy the amount allocated to the checking account by subjects in treatment D goes down by about 628 Euros, which means in turn that those subjects with less knowledge and experience in investment issues are significantly deterred from investing into the funds at all if the information on the risk and reward profiles is missing. The result on risk tolerance is somewhat inconclusive. We would expect that throughout all treatments a lower degree of risk aversion (or higher degree of risk appetite) decreases the amount on the checking account, as more money is allocated to the risky investment options. In fact, the allocations to the single funds that we will discuss in more detail further below are significantly driven by risk preferences, regardless of experimental treatments. However, while the conditional main effect of risk tolerance on the amount left on the checking account is insignificant, only in treatment C subjects behave according to our prediction.

---

<sup>51</sup> The results of the likelihood ratio tests are as follows: the coefficients for treatments B and C are not significantly different with  $F(1, 42) = 0.45$  and  $p = 0.5014$ , coefficients for treatments B and D are not significantly different with  $F(1, 142) = 0.01$  and  $p = 0.9309$ , and coefficients for treatments C and D are not significantly different with  $F(1, 142) = 0.29$  and  $p = 0.5934$ .

Table 3.11: Allocations to Checking Account

Checking Account	$\beta$	Robust SE	t-value	p-value
Constant	937.633***	360.896	2.60	0.010
Treatment B	81.858	484.702	0.17	0.866
Treatment C	-204.165	511.968	-0.40	0.691
Treatment D	44.178	482.971	0.09	0.927
Age (mc)	0.266	12.885	0.02	0.984
Female	397.118	472.299	0.84	0.402
Professional	-142.479	391.005	-0.36	0.716
Financial Literacy (mc)	-145.905	232.551	-0.63	0.531
Risk Tolerance (mc)	-31.582	213.047	-0.15	0.882
Comparability Index (mc)	-533.995**	242.552	-2.20	0.029
Need for Information and Consulting Index (mc)	-196.752	196.656	-1.00	0.319
Treatment B * Financial Literacy (mc)	155.717	316.926	0.49	0.624
Treatment C * Financial Literacy (mc)	586.310	364.909	-1.61	0.110
Treatment D * Financial Literacy (mc)	-628.454*	337.624	-1.86	0.065
Treatment B * Risk Tolerance (mc)	-244.900	313.287	-0.78	0.436
Treatment C * Risk Tolerance (mc)	-843.839**	412.379	-2.05	0.043
Treatment D * Risk Tolerance (mc)	312.747	414.059	-0.76	0.451

Note: Number of observations = 159. Standard errors are robust against heteroskedasticity. \*\*\*, \*\*, and \* indicate statistical significance at the 99%, 95%, and 90% confidence levels.  $F(16, 142) = 2.41$ ,  $p = 0.0032$ ,  $R^2 = 0.2914$ .

Taking another look at Table 3.10, the allocations to the four funds also differ between treatments. Fund BLUE receives equally high allocations in treatments A and D, while the invested amount in treatment B is clearly the lowest. The amount in treatment C is not significantly different from any of the others. Fund GREEN has the largest allocation in treatment C, but also the investment in treatment A is not significantly lower. On the other hand, in both treatments lacking the risk scale (B and D), much lower amounts are invested in the fund. Fund GREEN has an SRRI value of 3, which makes the fund the investment alternative which conforms best to the average risk tolerance of the test persons in the full

sample, at least if one assumes that the scale that we use for the self-assessment is equivalent to the SRRI scale.<sup>52</sup> This may explain why the allocations to Fund GREEN in treatments A and C are among the highest throughout the entire experiment. As for Fund RED, there are no statistically significant differences between any two treatments. The allocation to Fund WHITE in treatment B is significantly larger than in any other treatment, and there is no significant difference between the other three treatments. This is in line with the results on risk perception in treatment B.

Analogous to the multivariate analysis of the allocation to the checking account, we also regress each fund investment on the same selection of control variables to understand what drives the subjects' decisions. For the funds, we include an additional dummy variable in the regression that indicates whether the rank that each test person assigns to the riskiness of the fund in question is correct. All the regression models explain about 25 percent of the total variance, except for the model for Fund GREEN, which only explains only 12 percent. This might be a result of our conjecture that investments in Fund GREEN are most popular because they fit best to most people's risk tolerance. Individual differences in risk tolerance can only explain to which degree the other three funds are added to the portfolio, which is also reflected in the resulting regression coefficients of the *risk tolerance (mc)* variable in the respective models (negative and significant for Fund BLUE, positive and significant for funds RED and WHITE), but in the case of Fund GREEN, there might simply be too little variance.

---

<sup>52</sup> Again, there is no technical interrelation between the two scales, but it may be an obvious heuristic procedure for the test persons to match the SRRI of their preferred investment alternative with their risk tolerance value and vice versa.

Table 3.12: Allocations to Fund BLUE

Fund BLUE	$\beta$	Robust SE	t-value	p-value
Constant	2753.182***	559.7714	4.92	0.000
Treatment B	-985.305**	491.752	-2.00	0.047
Treatment C	-1171.559**	495.324	-2.37	0.019
Treatment D	335.850	635.822	0.53	0.598
Age (mc)	-22.676**	9.358	-2.42	0.017
Female	-549.244	405.508	-1.35	0.178
Professional	-1010.756***	368.642	-2.74	0.007
Financial Literacy (mc)	-115.943	220.756	-0.53	0.600
Risk Tolerance (mc)	-240.683*	139.535	-1.72	0.087
Risk Assessment Fund BLUE correct	794.769**	357.136	2.23	0.028
Comparability Index (mc)	144.562	196.321	0.74	0.463
Need for Information and Consulting Index (mc)	689.210***	251.474	2.74	0.007
Treatment B * Financial Literacy (mc)	172.173	243.931	0.71	0.481
Treatment C * Financial Literacy (mc)	8.313	270.203	0.03	0.976
Treatment D * Financial Literacy (mc)	559.042*	306.073	1.83	0.070
Treatment B * Need for Information and Consulting Index (mc)	-959.898***	347.498	-2.76	0.007
Treatment C * Need for Information and Consulting Index (mc)	-584.211	370.551	-1.58	0.117
Treatment D * Need for Information and Consulting Index (mc)	-583.138	469.719	-1.24	0.216

Note: Number of observations = 159. Standard errors are robust against heteroskedasticity. \*\*\*, \*\*, and \* indicate statistical significance at the 99%, 95%, and 90% confidence levels.  $F(17, 141) = 4.56$ ,  $p < 0.0001$ ,  $R^2 = 0.2664$ .

The observed conditional treatment main effects for Fund BLUE show that the differences in the investment amounts are actually caused by the experimental variations of the risk and reward section. Furthermore, pairwise comparisons using likelihood ratio tests reveal that the coefficients for *treatments B* and *D* ( $F(1, 141) = 6.08$ ,  $p = 0.0149$ ) and *treatments C* and *D* ( $F(1, 141) = 6.54$ ,  $p = 0.0116$ ) are significantly different from each other. There is no

significant difference between *treatments B* and *C* ( $F(1, 141) = 0.19, p = 0.6676$ ). Thus, subjects that either have the full section on the risk and reward profile available and those who have no explicit information at all invest significantly more than the participants that only have either one of the two constituent parts. Professionally experienced test persons invest significantly less in Fund BLUE than others. We do not know why this is the case, but it might have to do with for example the consideration of inflation-adjusted, returns that are not very attractive for Fund BLUE. Subjects with a higher risk tolerance also invest less throughout all treatments, as Fund BLUE does not fit their preferences properly. Interestingly, people who ranked Fund BLUE correctly as the least risky investment option invest significantly more than those who did not. This shows once again the importance for investors to be enabled to estimate investment risk correctly through appropriate documentation, especially at the edges of the spectrum: the least risky alternative is clearly attractive, and if investors have difficulties to detect that alternative, they tend to make adverse decisions. We see a similar effect for Fund WHITE which we will discuss further below. Those test persons who felt more uninformed and had a higher need for financial advice also invested more in Fund BLUE. Apparently, a less risky investment is preferable when making a relatively uninformed choice. Interestingly, this is not the case in treatment B, where the negative interaction effect neutralizes the conditional main effect. As it was the case for the checking account, financial literacy plays a role in treatment D only, in a way that more literate people invest more in Fund BLUE. We interpret this as the mirroring effect to lower investments in the checking account. The money that is not deposited on the account gets rather invested in a fund that bears only very little risk, but at least has a small return prospect. Finally, there is also a significant negative age effect, but its economic impact is fairly small.

Analyzing investments in Fund GREEN, we start with the resulting coefficients for the treatment dummies again. Participants in treatment D invest much less than in treatments A (see t-statistic in the regression table) and C (the likelihood ratio test yields  $F(1, 144) = 5.12$  and  $p = 0.0251$ ). The coefficient for *treatment B* lies in the middle and is not significantly different from any of the others (*B* vs. *C* yields  $F(1, 144) = 2.17$  and  $p = 0.1431$ ; *B* vs. *D* yields  $F(1, 144) = 0.33$  and  $p = 0.5654$ ). Subjects who state that they are better able to compare the investable funds allocate more money to Fund GREEN except for those who participated in treatment C.



Table 3.13: Allocations to Fund GREEN

Fund GREEN	$\beta$	Robust SE	t-value	p-value
Constant	2842.307***	664.581	4.28	0.000
Treatment B	-779.716	628.674	-1.24	0.217
Treatment C	183.022	597.369	0.31	0.760
Treatment D	-1164.196**	567.516	-2.05	0.042
Age (mc)	3.056	14.887	0.21	0.838
Female	-4.710	561.930	-0.01	0.993
Professional	113.784	523.296	0.22	0.828
Financial Literacy (mc)	-145.641	182.586	-0.80	0.426
Risk Tolerance (mc)	-100.400	157.837	-0.64	0.526
Risk Assessment Fund GREEN correct	329.482	427.527	0.77	0.442
Comparability Index (mc)	1047.938***	319.684	3.28	0.001
Need for Information and Consulting Index (mc)	88.230	181.903	0.49	0.628
Treatment B * Comparability Index (mc)	-731.110	635.213	-1.15	0.252
Treatment C * Comparability Index (mc)	-928.419*	552.182	-1.68	0.095
Treatment D * Comparability Index (mc)	-674.930	476.002	-1.42	0.158

Note: Number of observations = 159. Standard errors are robust against heteroskedasticity. \*\*\*, \*\*, and \* indicate statistical significance at the 99%, 95%, and 90% confidence levels.  $F(14, 144) = 2.12$ ,  $p = 0.0138$ ,  $R^2 = 0.1162$ .

As seen before in the univariate summary, there is no significant difference between the four treatments when it comes to investments in Fund RED. The multivariate analysis confirms this notion. None of the treatment dummy coefficients is significant and pairwise likelihood ratio tests are all insignificant, too.<sup>53</sup> Higher risk tolerance significantly increases the amount invested in Fund RED across all treatments. Furthermore, a higher evaluation of the comparability of the funds generally increases investments. Financial literacy also increases investments, but only in treatment B. Finally, a higher need for financial advice and further

<sup>53</sup> Treatment B vs. Treatment C:  $F(1, 141) = 0.44$ ,  $p = 0.5061$ . Treatment B vs. Treatment D:  $F(1, 141) = 0.19$ ,  $p = 0.6644$ . Treatment C vs. Treatment D:  $F(1, 141) = 1.64$ ,  $p = 0.2029$ .

information reduces the willingness to invest in Fund RED, but this is only true for treatment A, notably the experimental condition in which most information is provided. The interactions with the treatment dummies show that in all other treatments the effect is neutralized, even though the coefficient for the interaction with *treatment D* is only on the very edge of significance.

Table 3.14: Allocations to Fund RED

Fund RED	$\beta$	Robust SE	t-value	p-value
Constant	2101.762***	464.731	4.52	0.000
Treatment B	186.613	547.856	0.34	0.734
Treatment C	561.302	454.525	1.23	0.219
Treatment D	-76.461	482.994	-0.16	0.874
Age (mc)	-0.540	10.311	-0.05	0.958
Female	-238.133	439.203	-0.54	0.589
Professional	468.651	405.954	1.15	0.250
Financial Literacy (mc)	-173.459	148.208	-1.17	0.244
Risk Tolerance (mc)	333.796***	122.48	2.73	0.007
Risk Assessment Fund RED correct	-416.571	330.835	-1.26	0.210
Comparability Index (mc)	393.609**	183.268	2.15	0.033
Need for Information and Consulting Index (mc)	-499.860**	205.131	-2.44	0.016
Treatment B * Financial Literacy (mc)	482.241**	238.786	2.02	0.045
Treatment C * Financial Literacy (mc)	266.170	229.578	1.16	0.248
Treatment D * Financial Literacy (mc)	77.872	199.452	0.39	0.697
Treatment B * Need for Information and Consulting Index (mc)	793.378*	440.129	1.80	0.074
Treatment C * Need for Information and Consulting Index (mc)	529.943*	314.579	1.68	0.094
Treatment D * Need for Information and Consulting Index (mc)	526.153	319.729	1.65	0.102

Note: Number of observations = 159. Standard errors are robust against heteroskedasticity. \*\*\*, \*\*, and \* indicate statistical significance at the 99%, 95%, and 90% confidence levels.  $F(17, 141) = 4.54$ ,  $p < 0.0001$ ,  $R^2 = 0.2309$ .

Investments in Fund WHITE are much higher in treatment B than in treatment A (see Table 3.15) and treatment D (result of the likelihood ratio test is  $F(1, 144) = 4.92$ ,  $p = 0.0281$ ). The coefficients for the dummies *treatment B* and *C* are not significantly different from each other ( $F(1, 144) = 2.46$ ,  $p = 0.1190$ ), neither are *treatments C* and *D* ( $F(1, 144) = 0.08$ ,  $p = 0.7812$ ).

Table 3.15: Allocations to Fund WHITE

Fund WHITE	$\beta$	Robust SE	t-value	p-value
Constant	879.751	749.329	1.17	0.242
Treatment B	2566.353***	869.629	2.95	0.004
Treatment C	842.351	1061.761	0.79	0.429
Treatment D	556.447	846.997	0.66	0.512
Age (mc)	19.840	12.950	1.53	0.128
Female	254.864	582.031	0.44	0.662
Professional	611.120	470.375	1.30	0.196
Financial Literacy (mc)	123.692	158.708	0.78	0.437
Risk Tolerance (mc)	405.281***	153.993	2.63	0.009
Risk Assessment Fund WHITE correct	-19.122	658.916	-0.03	0.977
Comparability Index (mc)	-423.818*	241.292	-1.76	0.081
Need for Information and Consulting Index (mc)	-21.164	170.553	-0.12	0.901
Treatment B * Risk Assessment Fund WHITE correct	-1816.289*	1068.366	-1.70	0.091
Treatment C * Risk Assessment Fund WHITE correct	-616.469	1179.000	-0.52	0.602
Treatment D * Risk Assessment Fund WHITE correct	-185.858	968.457	-0.19	0.848

Note: Number of observations = 159. Standard errors are robust against heteroskedasticity. \*\*\*, \*\*, and \* indicate statistical significance at the 99%, 95%, and 90% confidence levels.  $F(14, 144) = 2.77$ ,  $p = 0.0012$ ,  $R^2 = 0.2427$ .

*Risk tolerance* has a significant positive effect on investments in Fund WHITE, while those subjects who are better able to compare the different funds invest less. This is an opposing effect to Funds RED and GREEN, where higher comparability leads to higher investments. Seemingly, the test persons shift funds away from the most risky investment opportunity if

they are better able to distinguish the available options. This interpretation is further supported by the fact that the interacted effect of subjects being in treatment B in which the risk assessment of Fund WHITE is biased downwards, but nevertheless having ranked Fund WHITE as the most risky alternative, is significantly negative: those subjects who identified the fund risk correctly invest less.

Overall, we conclude that in treatment B there are significantly lower investments in Fund BLUE, but significantly higher investments in Fund WHITE. This resembles the observed pattern in the risk ranking task that we discussed in the previous subsection, where the risk of Fund BLUE is overestimated and Fund WHITE is underestimated when the SRRI is taken away.

Taking another look back to Table 3.10, we see that the average SRRI value of the chosen allocations is the lowest in treatment D and the highest in treatment B.<sup>54</sup> The regression presented in Table 3.16 shows that this is indeed driven by the experimental variation. The difference between treatments A and D is insignificant, i.e. we conclude that there is no effect of introducing or removing the full section on the risk and reward profile in the KIDs. However, subjects in treatment B make a portfolio allocation that has a significantly higher average risk (a likelihood ratio test for difference between the resulting coefficients for *treatments B and D* yields significance with  $F(1, 147) = 8.18, p = 0.0049$ ), which is of course due to different amounts invested in Funds BLUE and WHITE. By integrating further control variables into the analysis, the difference between treatments A and C also becomes significant, and so does the difference between treatments C and D (result of the likelihood ratio test is  $F(1, 147) = 3.85, p = 0.0516$ ). There is no significant effect from varying the content of the KIDs when comparing treatments B and C (the likelihood ratio test yields  $F(1, 147) = 0.99, p = 0.3224$ ). Older people and test persons with relevant professional or educational backgrounds tend to make riskier portfolio allocations and of course people with a higher risk tolerance do the same.

---

<sup>54</sup> The average SRRI value does not reflect the actual portfolio risk accurately, because it neglects correlations between fund returns. We assume an SRRI value of 0 for the checking account.

Table 3.16: Value Weighted SRRI of Allocation

Value Weighted SRRI of Allocation	$\beta$	Robust SE	t-value	p-value
Constant	3.151***	0.372	8.48	0.000
Treatment B	0.770**	0.320	2.40	0.018
Treatment C	0.464*	0.271	1.72	0.088
Treatment D	-0.146	0.315	-0.46	0.644
Age (mc)	0.014*	0.008	1.85	0.067
Female	0.171	0.317	0.54	0.591
Professional	0.542**	0.267	2.03	0.044
Financial Literacy (mc)	0.074	0.091	0.82	0.414
Risk Tolerance (mc)	0.410***	0.100	4.12	0.000
$\rho_i^{risk}$	-0.413	0.262	-1.58	0.117
Comparability Index (mc)	0.069	0.143	0.48	0.628
Need for Information and Consulting Index (mc)	0.013	0.107	0.12	0.901

Note: Number of observations = 159. Standard errors are robust against heteroskedasticity. \*\*\*, \*\*, and \* indicate statistical significance at the 99%, 95%, and 90% confidence levels.  $F(11, 147) = 7.09$ ,  $p < 0.0001$ ,  $R^2 = 0.3311$ . The results hold when using TOBIT regression instead, as only 11 out of 159 observations are located on the limits 0 and 7.

We know from Table 3.5 that there are no significant differences between the average risk tolerance levels across treatments. Therefore, the observation of the different risk and return characteristics between the treatments raises the question under which condition the portfolio allocation fits better to the investors' preferences. We have seen that the risk rankings are significantly more accurate when subjects have the scale at hand, i.e. in treatments A and C, therefore the conjecture that the subjects in these treatments also show a higher degree of consistency between their risk tolerance and investment decisions seems plausible. The difficulty here is that the scale that we use to assess the subjects' risk tolerance is not equivalent to the SRRI measure. Both range from 1 to 7, with higher numbers indicating higher risk and risk tolerance, respectively, but it is not clear whether a subject with a risk tolerance of 7 should also choose an average portfolio risk of 7, i.e. should solely invest in Fund WHITE. In Table 3.10 we see in the last rows that the average difference between the SRRIs of the portfolios and their investors' risk tolerance level is very close to 0 (and also not significantly different from 0, as indicated by a Wilcoxon Signed Rank Test for the difference of the observed treatment median to the hypothetical value 0) in treatment C where only the

SRRI value is available in the KID. Therefore it seems that in this case the test persons simply “equalize” both scales, which makes a difference of 0 a natural benchmark level to test the single treatment medians against. In both treatments, B and C, this difference is not statistically significant, but it is clearly negative for treatments A and D. On the other hand, the median of this difference for the full sample is -0.3, i.e. the self-reported risk tolerance is generally higher than the average SRRI value of the portfolios, and there are good reasons why this is normatively appealing, too: considering an individual with the highest possible risk tolerance once again, there may be other reasonable motives and preferences at work that lower the resulting average SRRI of the chosen portfolio, for example the investor might want to diversify or has a preference for holding cash on the checking account for immediate availability. Therefore it may also be appropriate to test the treatment medians against the full sample median to see where the portfolio allocations deviate strongly from the individual risk tolerance levels. This is the case in treatment B, where the average portfolio risk is significantly higher, and treatment D, where it is significantly lower. However, due to the discussed measurement problems and the inconclusive results of our analyses, we refrain from presenting a final result on the fit between preferences and choices. This rather underpins the importance of individual clarification of this issue by each investor, i.e. to transform the provided information about the risk and reward profile of an investment option into an individual criterion whether or not to add it to one’s portfolio. Personal experience, as well as individual financial advice, may be helpful in this matter. A study on long-term effects of the standardization of information in repeated investment decisions may provide valuable insights and is a potentially fruitful field for further research.

### 3.3.5 Relevance of Fund Characteristics and Suggestions for Improvement of KIDs

In the investment stage, right after the actual portfolio allocation, we asked the participants to evaluate the importance of several fund characteristics for their investment decisions. Most of the attributes are clearly related to an item in the KIDs, namely investment policy, performance, volatility (not as a figure, but implicitly in the SRRIs and the performance charts), the risk and reward profile (verbal description and SRRI value), and costs, while others are not included in the documents, e.g. asset preservation and the fund portfolio structure. Furthermore, the test persons could name up to two additional factors that were important for their decision. Each attribute is evaluated on a scale from 1 (“very important”) to 5 (“not important”). Table 3.17 summarizes the results.

We determine differences between treatments by pairwise Kolmogorov-Smirnov-Tests. For the items that relate to specific KID elements, the overall picture is consistent with the results from the KID assessment stage presented in Table 3.6: the only significant difference that we observe is the higher importance of the investment policy in treatment B compared to treatment A.

Table 3.17: Importance of Fund Characteristics per Treatment

	Average per Treatment			
	A	B	C	D
<b>Importance of Investment Policy (KS-Test)</b>	2.540 <sub>a</sub>	1.830 <sub>b</sub>	2.110 <sub>a,b</sub>	2.190 <sub>a,b</sub>
<b>Importance of Performance (KS-Test)</b>	2.230 <sub>a</sub>	2.230 <sub>a</sub>	2.270 <sub>a</sub>	2.580 <sub>a</sub>
<b>Importance of Volatility (KS-Test)</b>	2.530 <sub>a</sub>	2.440 <sub>a</sub>	2.410 <sub>a</sub>	2.830 <sub>a</sub>
<b>Importance of Asset Preservation (KS-Test)</b>	2.210 <sub>a</sub>	1.920 <sub>a</sub>	2.190 <sub>a</sub>	2.190 <sub>a</sub>
<b>Importance of Verbal Description of Risk and Reward Profile (KS-Test)</b>	3.130 <sub>a</sub>	2.770 <sub>a</sub>		
<b>Importance of Costs (KS-Test)</b>	2.550 <sub>a</sub>	2.670 <sub>a</sub>	2.670 <sub>a</sub>	2.590 <sub>a</sub>
<b>Importance of SRRI Values (KS-Test)</b>	2.410 <sub>a</sub>		2.220 <sub>a</sub>	
<b>Importance of Portfolio Structure (KS-Test)</b>	2.440 <sub>a</sub>	2.030 <sub>a</sub>	2.710 <sub>a</sub>	2.560 <sub>a</sub>

Note: The items “Importance of Verbal Description of Risk and Reward Profile” and “Importance of SRRI Values” were not available in all treatments. The indicated means for “Importance of...” variables do not include test persons who answered “don’t know / no statement”. Means which do not share the same subscript are significantly different from each other as indicated by pairwise Kolmogorov-Smirnov-Tests with a confidence level of 90 percent or higher. Detailed results of pairwise comparisons are available from the authors upon request. Additional descriptive statistics can be found in Appendix 3.J.

As for the description of the funds’ risk and reward profiles, we observe that the importance of the SRRI, if available, is generally ranked higher than the importance of the verbal descriptions. However, concluding that verbal descriptions should be abandoned from the KIDs would be misleading. On the contrary, in order to make the SRRI more transparent, we suggest that a volatility figure should also explicitly be reported in the KID, as it is the basis

of the SRRI calculations and can be helpful to better understand the differences between the seven classifications.<sup>55</sup>

In order to improve the comparability of costs, we propose to include an additional scale that, in analogy to the SRRI, compares the costs of an investment fund with those of other investment products. There are remarkable differences in fee structures across asset classes and especially unexperienced investors might benefit a lot from an increase in transparency in this matter.

Asset preservation is a fund property and also a personal motive that is regarded as very important compared to most of the other listed characteristics over all treatments. Therefore, another suggestion to improve the usability of KIDs for investors is to include an explicit statement whether there is some form of capital protection or guarantee for a fund or not. The funds' portfolio structure is regarded as comparably important to the other fund characteristics. While the statements on the objectives and investment policy provide a general overview over the target universe of a fund, it may therefore be interesting to provide for example a pie chart that depicts the targeted strategic asset allocation. If a fund uses derivatives for hedging purposes, they should also be included in such an overview to help the investors understand the relative size (and importance) of these positions.

The participants are given the opportunity to name up to two further fund properties that are important to them when making their investment decision. Over all treatments, 75 entries were made by 53 test persons. It seems that most test persons report other determinants of their decision, mainly personal preferences, expectations, and experiences, instead of actual fund properties. In total, 40 of the given answers can be categorized as personal or subjective determinants. Furthermore, 11 entries named a fund property that is already mentioned in the questionnaire, for example costs or past performance. There are 12 subjects who mention the funds' benchmark or their relative performance as an important criterion for their decision.<sup>56</sup> The rest of the given answers loosely addresses other fund properties or personal attitudes, like for example a lack of social and ethical agreeableness of the funds, the depository bank being incorporated in Luxembourg, or the unwillingness to support financial speculations. A full list of these additional items is available from the authors upon request.

---

<sup>55</sup> Also, picking up the example of the MCC that we discussed in the introduction of this chapter, it could be mentioned that the SRRI is calculated according to binding regulations that are the same for all UCITS.

<sup>56</sup> Benchmarks and benchmark returns are, if applicable, included in KIDs in the sections "Objectives and Investment Policy" and "Past Performance". All funds in our experiment have a benchmark.



While we are confident that our list of fund properties is largely comprehensive in capturing the most important items, the examination of freely given answers shows that of course personal determinants like preferences, investment goals, expectations, and values play an important role in investment decisions. While KIDs strive to make information about investment products accessible and comparable to potential investors, they do not help investors to define these personal attitudes and find the most suitable portfolio of investment products. For example, using KIDs is helpful to compare the costs of funds, but the investor still has to determine her willingness to pay.

### **3.4 Discussion**

In this chapter we present the results of an experimental study on the variation of descriptions of the risk and reward profile of investment funds in Key Investor Information Documents. We use plain investment funds with risk and reward profiles that spread across the entire range of the SRRI scale, i.e. the available options differ heavily in their risk and reward profiles. This makes the investment decision much simpler for the test persons compared to a situation where they can choose between different fund structures (remember that KIDs are also mandatory for other UCITS like REITs, ETFs, funds of funds, capital protection funds etc.) and between investment options that are not easily distinguishable. The importance of the risk and reward profile on the whole may be much higher and therefore the treatment effect of removing it from the KIDs entirely or in parts could be more severe if the subjects have to make such more demanding investment decisions.

We find that the removal of the entire section, which consists of a numerical scale and a verbal description of risk factors diminishes the ability to estimate the risk of the available investment options. However, there is no difference in the perception of the test person's ability to compare a range of fund characteristics. Especially for people with little financial knowledge and experience there is a tendency towards lower total investments when the information is missing while for all others the propensity to invest is not influenced. Through shifts between the funds, the resulting average SRRI value of the asset allocation is not significantly different from the portfolio that is chosen by subjects who have full information.

We also observe significant changes in investment behavior and risk perception when only a verbal description of risk factors is provided. Risk estimates are even more inconsistent than in treatment D where no information is given at all and portfolio risk increases considerably. Of course one should not conclude to abandon the given information from KIDs, even though

this would have a principally desirable effect on risk assessment and portfolio choice. Instead we argue that information on a fund property that is presented verbally should be accompanied by a meaningful numerical scale whenever possible, as this helps the investor to assess the relative importance of the property in question and whether it is favorable to her or not. The interpretation of verbal content is more accurate when the scale is available.

This point is especially important, as the SRRI is a unique feature of KIDs. Regulation requires fact sheets (“Produktinformationsblatt”, PIB in Germany)<sup>57</sup> that contain verbal descriptions of the risk and reward profiles only for most other investment products on the market, e.g. for stocks, bonds, certificates, options, futures, and savings products. Taking into account our results, PIBs should be adapted to KID standards by including an SRRI scale whenever possible to avoid misinterpretations. In the previous section we have already discussed possible extensions of KIDs themselves, like for example cost scales. To test these extensions and especially to develop informative scale measures for other properties of investment funds (and other products) is another field for further research.

Finally, when subjects only have the scale at hand, their ability to estimate fund risk correctly remains unchanged, but they tend to invest more riskily. However, it is unclear whether the allocation in treatment A or C actually fits better to the individual preferences.

The description of the funds’ objectives and investment policies is a very important feature, yet it is relatively difficult to make comparisons between different funds, as Table 3.6 shows. Further opportunities for standardization of this section are certainly very limited. However, one could point out specific attributes of a fund in a portfolio context, like for example diversification potential with regard to geographical, industrial, currency, or maturity focusing. Following a recent court decision in Germany (see OLG Stuttgart (2013)), a private investor received an indemnity sum from her bank due to a case of miscounseling. In the justification, the judges argue that the investor could not understand the roles of single positions within her portfolio with regards to their effects on total risk. The bank used verbal classifications like for example “growth” and “chance” to categorize investments by their typical risk and reward profile. While such a categorization appears somewhat more subjective than the SRRI applied in KIDs, the court decision stresses the importance of a more holistic perspective on risk and reward properties. Furthermore, it becomes clear that

---

<sup>57</sup> An overview of PIBs for various types of certificates, for example, is available on the website of the German Derivatives Association (DDV): <http://www.derivateverband.de/DEU/Transparenz/Produktinformationsblaetter>

standardization and categorization gives rise to negligence of details that are specific to the personal situation, experience, and preferences of the investor. Investors (and financial advisors) need to learn how to incorporate the information provided in KIDs and PIBs into their decision making criteria in order to select the investment options that best fit their needs. With regards to the risk and reward profile, an interesting approach would be for banks and other financial advisors to measure their clients' willingness and ability to take on financial risks in a way that it is easily transferable to the SRRI scale. For example, the volatility of a client portfolio could be varied in repeated simulations to determine a critical upper value that in turn could easily be categorized according to the thresholds of the SRRI scale.

Supporting private investors by making information about financial products easily accessible and comparable is surely desirable and therefore we expect ongoing initiative in this field. However, to clarify this once more, reducing manifold individual preferences on the one hand and time-varying, unpredictable risk and reward properties of complex investment products on the other hand, to a single number may be appealing to make complex issues more tractable. But it should not be forgotten that all complexities are in fact still there.

## **Appendix to Chapter 3**

### **Appendix 3.A / Introduction and Welcome Screen**

Thank you for your interest in the scientific experiment “Investment Decisions and Product Information”!

The experiment is conducted by the Chair of Empirical Capital Market Research at WHU – Otto Beisheim School of Management in Vallendar and the collected data will be used by Maximilian Troßbach in his doctoral thesis. If you have questions or suggestions, please send me an email: [maximilian.trossbach@whu.edu](mailto:maximilian.trossbach@whu.edu)

Participation in the experiment is anonymous and no data will be collected that could be used to identify you in person. We will ask you to assess a fictitious financial investment and your general experience with private investments. There are no objectively right or wrong answers; we are rather interested in your personal perspectives and attitudes. Likewise, there is no specific previous knowledge required – your participation in the experiment is highly appreciated, no matter if you have no, little, or much experience in the field of private investments.

The experiment will take about half an hour.

In order to run the experiment on your computer, you need Acrobat Reader or comparable software to read pdf files. Most computers are already equipped with suitable software. You can also download Acrobat Reader [here](#).

If you are not sure if you can read pdf files on your computer, you can open a test document [here](#). If you cannot open and read the document, please install Acrobat Reader on your computer or use a different device that is capable of reading pdf files.

Participation in the experiment is voluntary. By clicking the “next” button below, you agree to participate and the experiment starts. You can stop the experiment at any time. Upon completing the experiment entirely, you can take part in a lottery where you can win one of 15 amazon gift coupons worth 20 euros each. On the last page of this online questionnaire, we will provide you with the keyword to take part in the lottery. By sending us the keyword via email, you automatically take part in the lottery. At the same time, we ensure that your answers in the online questionnaire are collected separate from your personal data (name,

email address) to ensure anonymity of your answers. The procedure of the lottery will be explained to you once again at the end of the questionnaire.

Please close all other windows of your internet browser and all other open applications now to remain undisturbed while answering the online questionnaire.

Next

## Appendix 3.B / Investment Stage

In the following we describe a situation to you that is typical in the context of an investment decision (for example for your private retirement provisions). Please look at all the information that is provided to you about the available investment opportunities. Please make your decision as if it was for real, as if it was your own money that you can invest.

Please refrain from searching any additional information (other than provided in the experiment) about the investment alternatives. We would like to investigate how your investment decisions depend on the available product information; therefore your data is unusable if you obtain any further external information.

[Back](#)[Next](#)

Imagine you have an amount of 10,000 Euros at your disposal. You can transfer the money to an interest-free checking account and invest it in four different investment funds that will be described in the following. You can choose to distribute the total amount to these five options any way you like.

Before you make your decision, please take your time to look at the following information. Behind the links below you find a document, the so-called Key Investor Information Document, for each of the funds. These documents are provided by the investment companies for existing and potential investors. They are information sheets that are prescribed by law and they describe selected characteristics of the particular investment product in a standardized manner. The investment funds are really existent. For the purpose of the experiment, only the names of the funds, of the investment company, and the depository bank, as well as the security identification numbers have been changed in a way that it is not possible for you to recognize the funds. All other details are true to original.

By clicking one of the links below a new window will be opened in your browser that displays the respective information document as a pdf file. Please open all four documents and leave them open during the entire course of the experiment in order to have access to all the information at any time. This is meant to facilitate your decision-making.

Fund BLUE

Fund GREEN

Fund RED

Fund WHITE

Please click “Next” only after you have looked at all the information unhurriedly. You will then immediately proceed to the submission of your investment decision.

Back

Next

Imagine you have an amount of 10,000 Euros at your disposal. You can transfer the money to an interest-free checking account and invest it in the four previously presented investment funds. You can choose any partition of the entire amount. Please enter your preferred amount in Euro for each of the available options.

- Please use integers (no decimal places)
- If you do not want to invest in specific funds or the checking account, please enter a “0” in each case
- If you would like to invest the entire amount in a specific fund or transfer it entirely to the checking account, please enter “10000” (no separator or space)
- Please notice that the sum of all investments plus the amount on the checking account has to be exactly 10000.

I would like to invest \_\_\_\_\_ Euros in Fund BLUE.

I would like to invest \_\_\_\_\_ Euros in Fund GREEN.

I would like to invest \_\_\_\_\_ Euros in Fund RED.

I would like to invest \_\_\_\_\_ Euros in Fund WHITE.

I would like to transfer \_\_\_\_\_ Euros to the interest-free checking account.

Back

Next

Please rate the characteristics below by their importance. We would like to know how important these fund characteristics were in your personal investment decision that you made in the previous section of the experiment.

	Very important	Important	Fairly important	Less important	Not important	Don't know / no statement
Investment policy of the funds	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Past financial performance of the funds	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Past value fluctuations of the funds	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Safety of the invested capital (asset preservation)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Verbal descriptions of the risk and return profiles of the funds	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Costs of the funds (one-time and recurring)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Risk classes of the funds according to the Key Investor Information Documents	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Portfolio structure of the funds (distribution of the fund capital)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
This additional fund characteristic has influenced my investment decision: _____	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
This additional fund characteristic has influenced my investment decision: _____	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Back

Next



We would like to know now how you rate the risk of the funds in a direct comparison. Please assign ranks from 1 to 4. The fund that is the least risky in your opinion shall be ranked 1. The fund that you rate the riskiest shall be ranked 4, and the two funds in between shall be ranked 2 and 3 analogously. Please assign each rank only once.

Fund BLUE \_\_\_\_\_

Fund GREEN \_\_\_\_\_

Fund RED \_\_\_\_\_

Fund WHITE \_\_\_\_\_

Back

Next

## Appendix 3.C / Self-Assessment Stage

In the following, we would like you to provide some information about you as a person and your experience in the field of private investments. This information is of high importance to us for the scientific usability of the collected data. It is not possible to identify you personally from the data.

Age \_\_\_\_\_ years

### Gender

☐ female      ☐ male

### What is your highest obtained school or academic degree?

- ☐ none
- ☐ Certificate of Secondary Education
- ☐ General Certificate of Secondary Education
- ☐ Advanced Technical College Entrance Qualification
- ☐ General Qualification for University Entrance
- ☐ Polytechnic Degree
- ☐ University Degree
- ☐ Doctorate Degree

Back

Next

Have you been involved with financial investments in your current or a former job or during your studies or any professional training?

☐ Yes      ☐ No

Back

Next

**In your opinion, how familiar are you with the topic “financial investments” (i.e. types of investments, financial products, financial markets etc.)?**

“1” indicates that you are not familiar with the topic at all. “4” indicates that you are as well versed as the population average. “7” indicates that you know as much as a professional investor.

	1	2	3	4	5	6	7	
not familiar at all	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	like a professional investor

Back

Next

**How do you rate your readiness to assume risk in financial investments?**

1 = I would like to assume as little risk as possible and therefore I am willing to forgo return opportunities

4 = I am ready to assume increased risk for higher return opportunities

7 = I am ready to assume very high risk including the risk of total loss in order to get very high return opportunities

	1	2	3	4	5	6	7	
As little risk as possible / low return opportunities	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	very high risk / very high return opportunities

Back

Next

## Appendix 3.D / KID Assessment Stage

Finally, we ask you to evaluate the Key Investor Information Documents that were available for your investment decision. Please indicate to what extent you agree to the statements below.

I was able to compare the objectives and investment policies of the funds without any problems.

<b>fully disagree</b>	<b>partly disagree</b>	<b>undecided</b>	<b>partly agree</b>	<b>fully agree</b>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

It was particularly important for my investment decision to compare the objectives and investment policies of the funds.

<b>fully disagree</b>	<b>partly disagree</b>	<b>undecided</b>	<b>partly agree</b>	<b>fully agree</b>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

I was able to compare the risk and reward profiles of the funds without any problems.

<b>fully disagree</b>	<b>partly disagree</b>	<b>undecided</b>	<b>partly agree</b>	<b>fully agree</b>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

It was particularly important for my investment decision to compare the risk and reward profiles of the funds.

<b>fully disagree</b>	<b>partly disagree</b>	<b>undecided</b>	<b>partly agree</b>	<b>fully agree</b>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

I was able to compare the past financial performances of the funds without any problems.

<b>fully disagree</b>	<b>partly disagree</b>	<b>undecided</b>	<b>partly agree</b>	<b>fully agree</b>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

It was particularly important for my investment decision to compare the past financial performances of the funds.

fully disagree	partly disagree	undecided	partly agree	fully agree
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

I was able to compare the costs of the funds without any problems.

fully disagree	partly disagree	undecided	partly agree	fully agree
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

It was particularly important for my investment decision to compare the costs of the funds.

fully disagree	partly disagree	undecided	partly agree	fully agree
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

I was able to compare the funds in general without any problems.

fully disagree	partly disagree	undecided	partly agree	fully agree
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

It was particularly important for my investment decision to compare the funds in general.

fully disagree	partly disagree	undecided	partly agree	fully agree
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Actually I would have needed to contact an investment advisor prior to my decision.

fully disagree	partly disagree	undecided	partly agree	fully agree
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

I would have liked to acquire further information about the available funds prior to my decision.

fully disagree	partly disagree	undecided	partly agree	fully agree
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Back

Next

### Appendix 3.E / Debriefing Screen

You have completed the experiment. Thank you very much for your participation!

If you like, you can take part in the lottery of **15 amazon coupons worth 20 Euros** each now. For this purpose, please note the keyword “investment funds” and send it to my email address [maximilian.trossbach@whu.edu](mailto:maximilian.trossbach@whu.edu).

The winners will receive a code via email that can be redeemed at the amazon online shop after the survey is completed.

You may close this window and all other windows that you have opened during the experiment now.

## Appendix 3.F / KID Fund BLUE<sup>58</sup>

### Key Investor Information

This document provides you with key investor information about this fund. It is not marketing material. The information is required by law to help you understand the nature and the risks of investing in this fund. You are advised to read it so you can make an informed decision about whether to invest.

### Fund BLUE

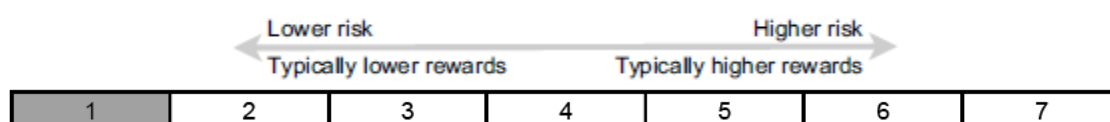
WKN: 123456 ISIN: LU001111111

Managed by XY Fund Company S.A.

### Objectives and Investment Policy

The objective of the investment policy is to yield a return in Euros that is geared to the benchmark (3M EUR LIBID). To achieve this, the fund invests into government and corporate bonds that are denominated or secured in Euros. The average fixed interest period is binding for at most 12 months and is managed by using appropriate derivatives and other measures. Derivatives can also be used for other portfolio management purposes. The selection of the single investments is at the discretion of the fund management. The fund is geared to a benchmark. It does not replicate the benchmark, but tries to outperform it and could therefore deviate considerably – positively and negatively – from the benchmark. The fund is subject to various risks. A detailed risk assessment and further remarks can be found in the prospectus in the chapter “risks”. The currency of Fund BLUE is EUR. Returns and capital gains are not paid out but reinvested in the fund. You can claim the redemption of fund units at each valuation date. Redemption may only be suspended in exceptional cases and in due consideration of your interest as an investor.

### Risk and Reward Profile



The calculation of the risk and reward profile is based on historical data that cannot be taken as a reliable basis for the future risk profile. This risk indicator is subject to changes; the classification of the fund may change over time and cannot be guaranteed. Even a fund that is classified the lowest category (category 1) is not a completely risk-free investment. This fund is classified as category 1, because unit prices usually hardly fluctuate and therefore chances of both, losses and gains, should be low. The following risks are essential to the fund and are not covered appropriately by the risk and reward profile:

The fund invests a significant share in bonds whose value depends on the ability of their issuers to make their payments. The risk of a shortfall in payment is always existent and may lead to your investment incurring a loss. The fund engages in derivative transactions with different contractual partners to a significant amount. In the case that a contractual partner does not make payments (for example in the case of insolvency) this may lead to your investment incurring a loss.

---

<sup>58</sup> The funds we use in our experiment are marketed in Germany and KIDs are not available in English. Therefore we present our own translations in this appendix. We used a template issued by the Committee of European Securities Regulators (CESR 2010a) when making these translations.

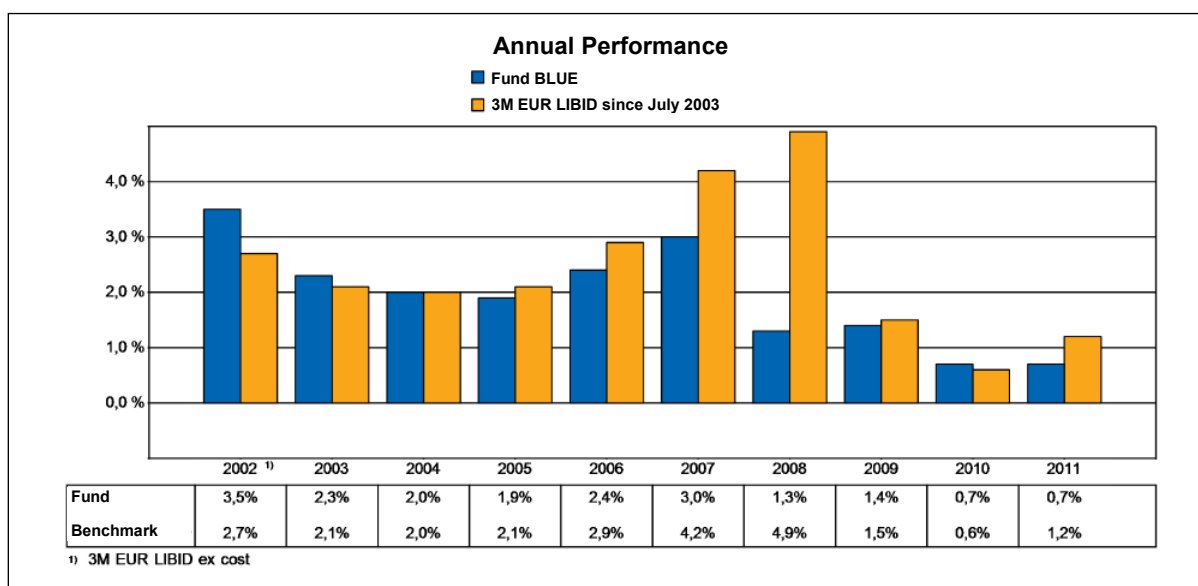
## Charges for this Fund

The charges you pay are used to pay the costs of running the fund, including the costs of marketing and distributing it. These charges reduce the potential growth of your investment.

One-time costs pre and past investing	
<b>Issue surcharges</b>	1.00%. This is the maximum amount that is deducted from your investment prior to investing
<b>Redemption fees</b>	No redemption fees
Costs that are deducted annually from the fund	
<b>Current Costs</b>	0.42%
Costs the fund has to bear under certain circumstances	
Fees tied to the fund's performance	None

The indicated issue surcharge is the maximum amount. It may be lower in a particular case. The amount in effect for you can be obtained from the department which is responsible for you or from your financial advisor. The reported current costs incurred in the business year of the fund that ended on December 31st, 2011. They may change year by year. Current costs do not include performance fees and transaction costs. Further information about costs can be found in the prospectus in the chapter "costs".

## Past Performance



The past performance is not a reliable indicator for future performance. All charges and fees that have been incurred by Fund BLUE were subtracted in the calculation of the performance figures. Issue surcharges and redemption fees are not considered. Fund BLUE came into existence in 1988. The performance is calculated in EUR.

## Practical Information

The depository is ABC Bank Luxembourg S.A., Luxembourg. The prospectus, annual and semi-annual reports are available in the language of this document, as well as in English at XY Fund Company, S.A. and can be requested at no charge. You can obtain the latest unit prices and further information regarding this fund in the language of this document as well as in English on the local website of XY Fund Company, S.A. The fund is subject to Luxembourgian tax law. This may have an effect on how your personal income from the fund is taxed. XY Fund Company, S.A. can be held liable solely on the basis of any statement that is contained in this document which is misleading, inaccurate, or inconsistent with the relevant parts of the prospectus for the fund. This fund is authorized in Luxembourg and regulated by the Commission de Surveillance du Secteur Financier.



## Appendix 3.G / KID Fund GREEN

### Key Investor Information

This document provides you with key investor information about this fund. It is not marketing material. The information is required by law to help you understand the nature and the risks of investing in this fund. You are advised to read it so you can make an informed decision about whether to invest.

### Fund GREEN

WKN: 654321 ISIN: LU000011111

Managed by XY Fund Company S.A.

### Objectives and Investment Policy

The objective of the investment policy is to sustainably generate additional capital gain compared to the benchmark (iBoxx Euro Overall 7-10Y). To achieve this, the fund invests into government bonds, bonds from close-state issuers and covered bonds. Corporate bonds and bonds from issuers in emerging markets can be added. Close-state issuers are for example central banks, government authorities, regional authorities and supra-national institutions. The average maturity of the investments is in the range between 7 to 10 years. The selection of the single investments is at the discretion of the fund management. The fund is geared to a benchmark. It does not replicate the benchmark, but tries to outperform it and could therefore deviate considerably – positively and negatively – from the benchmark. The fund is subject to various risks. A detailed risk assessment and further remarks can be found in the prospectus in the chapter “risks”. The currency of Fund GREEN is EUR. Returns and capital gains are not paid out but reinvested in the fund. You can claim the redemption of fund units at each valuation date. Redemption may only be suspended in exceptional cases and in due consideration of your interest as an investor.

### Risk and Reward Profile



The calculation of the risk and reward profile is based on historical data that cannot be taken as a reliable basis for the future risk profile. This risk indicator is subject to changes; the classification of the fund may change over time and cannot be guaranteed. Even a fund that is classified the lowest category (category 1) is not a completely risk-free investment. The fund is classified as category 3, because unit prices fluctuate relatively little and therefore chances of both, losses and gains, are relatively small. The following risks are essential to the fund and are not covered appropriately by the risk and reward profile:

The fund invests a significant share in bonds whose value depends on the ability of their issuers to make their payments. The risk of a shortfall in payment is always existent and may lead to your investment incurring a loss. The fund engages in derivative transactions with different contractual partners to a significant amount. In the case that a contractual partner does not make payments (for example in the case of insolvency) this may lead to your investment incurring a loss.

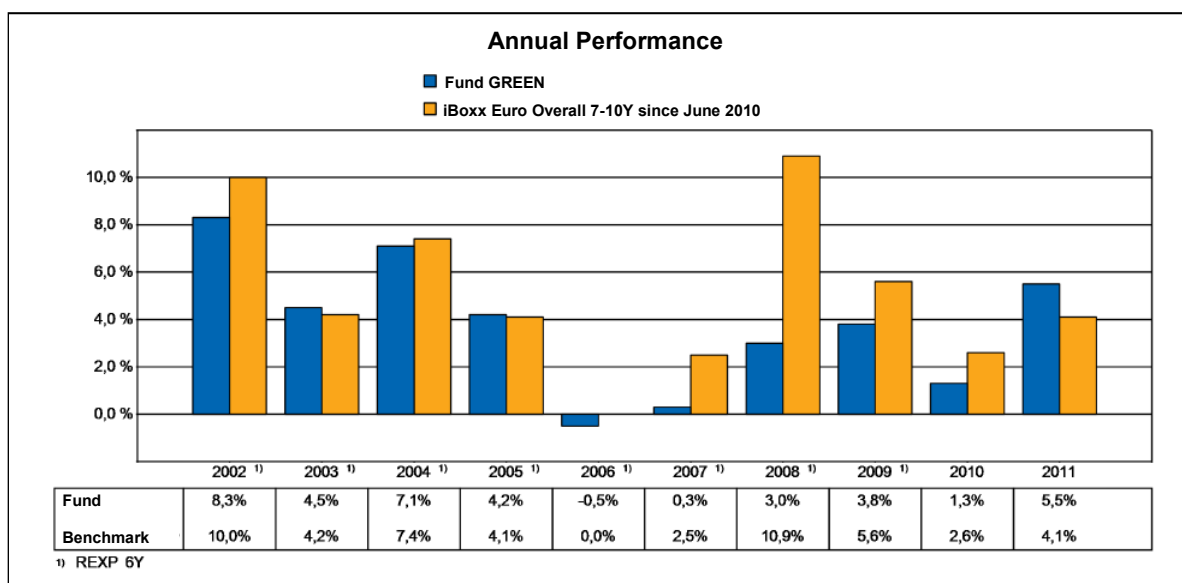
## Charges for this fund

The charges you pay are used to pay the costs of running the fund, including the costs of marketing and distributing it. These charges reduce the potential growth of your investment.

One-time costs pre and past investing	
Issue surcharges	3.00%. This is the maximum amount that is deducted from your investment prior to investing
Redemption fees	No redemption fees
Costs that are deducted annually from the fund	
Current Costs	0.80%
Costs the fund has to bear under certain circumstances	
Fees tied to the fund's performance	0.48% The performance-related pay corresponds to 25% of the amount by which the fund's performance is higher than the benchmark's performance. Details can be found in the prospectus in the chapter "Costs and Services".
Compensation from securities lending	0.14%

The indicated issue surcharge is the maximum amount. It may be lower in a particular case. The amount in effect for you can be obtained from the department which is responsible for you or from your financial advisor. The reported current costs incurred in the business year of the fund that ended on December 31st, 2011. They may change year by year. Current costs do not include performance fees and transaction costs. Further information about costs can be found in the prospectus in the chapter "costs".

## Past Performance



The past performance is not a reliable indicator for future performance. All charges and fees that have been incurred by Fund GREEN were subtracted in the calculation of the performance figures. Issue surcharges and redemption fees are not considered. Fund GREEN came into existence in 1992. The performance is calculated in EUR.

## Practical Information

The depository is ABC Bank Luxembourg S.A., Luxembourg. The prospectus, annual and semi-annual reports are available in the language of this document, as well as in English at XY Fund Company, S.A. and can be requested at no charge. You can obtain the latest unit prices and further information regarding this fund in the language of this document as well as in English on the local website of XY Fund Company, S.A. The fund is subject to Luxembourgian tax law. This may have an effect on how your personal income from the fund is taxed. XY Fund Company, S.A. can be held liable solely on the basis of any statement that is contained in this document which is misleading, inaccurate, or inconsistent with the relevant parts of the prospectus for the fund. This fund is authorized in Luxembourg and regulated by the Commission de Surveillance du Secteur Financier.

# Appendix 3.H / KID Fund RED

## Key Investor Information

This document provides you with key investor information about this fund. It is not marketing material. The information is required by law to help you understand the nature and the risks of investing in this fund. You are advised to read it so you can make an informed decision about whether to invest.

## Fund RED

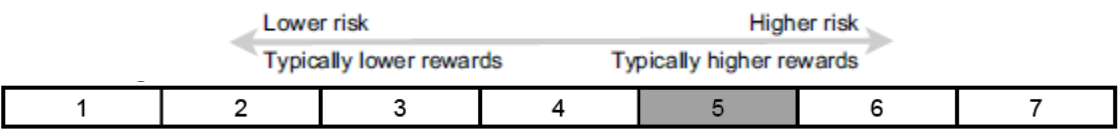
WKN: 333333 ISIN: LU000000011

Managed by XY Fund Company S.A.

## Objectives and Investment Policy

The objective of the investment policy is to sustainably generate additional capital gain compared to the benchmark (ML Euro BB-B Non-Financial Fixed & FRN High Yield Constrained). To achieve this, the fund invests in Euro-denominated corporate bonds, with the focus being on so-called high yield bonds. The selection of the single investments is at the discretion of the fund management. The fund is geared to a benchmark. It does not replicate the benchmark, but tries to outperform it and could therefore deviate considerably – positively and negatively – from the benchmark. The fund is subject to various risks. A detailed risk assessment and further remarks can be found in the prospectus in the chapter “risks”. The currency of Fund RED is EUR. Income from the fund is paid out yearly. You can claim the redemption of fund units at each valuation date. Redemption may only be suspended in exceptional cases and in due consideration of your interest as an investor.

## Risk and Reward Profile



The calculation of the risk and reward profile is based on historical data that cannot be taken as a reliable basis for the future risk profile. This risk indicator is subject to changes; the classification of the fund may change over time and cannot be guaranteed. Even a fund that is classified the lowest category (category 1) is not a completely risk-free investment. The fund is classified as category 5, because unit prices fluctuate relatively strong and therefore chances of both, losses and gains, are relatively high. The following risks are essential to the fund and are not covered appropriately by the risk and reward profile:

The fund invests a significant share in bonds whose value depends on the ability of their issuers to make their payments. The risk of a shortfall in payment is always existent and may lead to your investment incurring a loss. The fund engages in derivative transactions with different contractual partners to a significant amount. In the case that a contractual partner does not make payments (for example in the case of insolvency) this may lead to your investment incurring a loss. The fund invests a significant share of its assets into corporate bonds without investment grade. Focusing on this kind of corporate bonds bears the risk that issuers are facing economic difficulties with a higher probability and therefore are more likely to default. The value of these bonds can then go down to zero and negatively influence the fund’s assets directly.

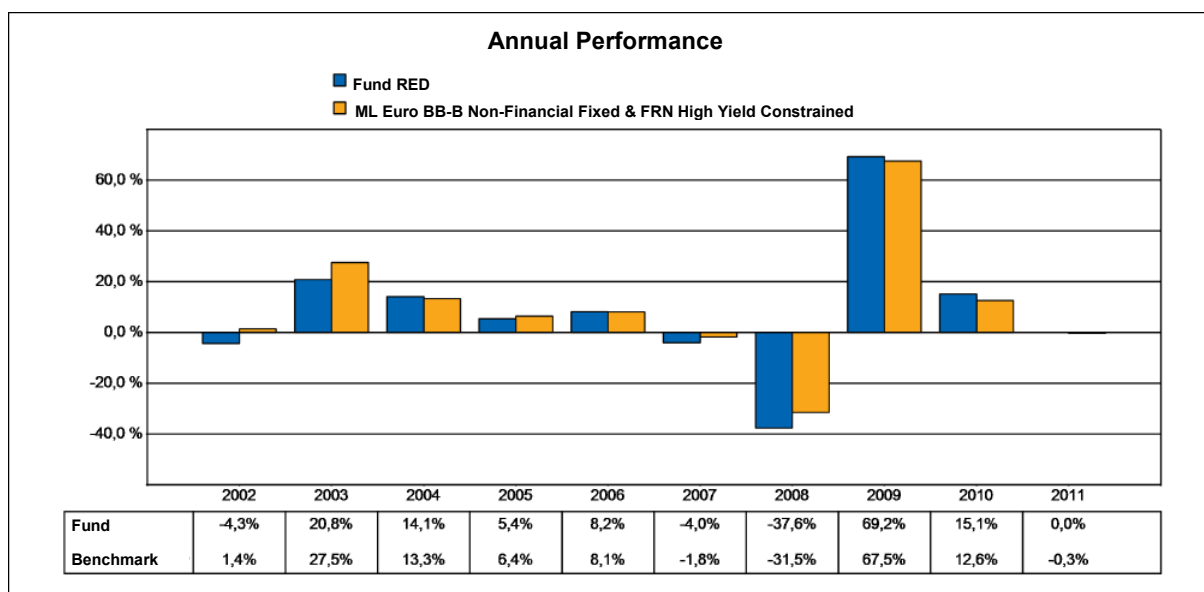
## Charges for this fund

The charges you pay are used to pay the costs of running the fund, including the costs of marketing and distributing it. These charges reduce the potential growth of your investment.

One-time costs pre and past investing	
Issue surcharges	3.00%. This is the maximum amount that is deducted from your investment prior to investing
Redemption fees	No redemption fees
Costs that are deducted annually from the fund	
Current Costs	1.14%
Costs the fund has to bear under certain circumstances	
Fees tied to the fund's performance	0.06% The performance-related pay corresponds to 25% of the amount by which the fund's performance is higher than the performance of the Merrill Lynch Euro BB-B Non-Financial Fixed & Floating Rate HY Constrained. Details can be found in the prospectus in the chapter "Costs and Services".
Compensation from securities lending	0.01%

The indicated issue surcharge is the maximum amount. It may be lower in a particular case. The amount in effect for you can be obtained from the department which is responsible for you or from your financial advisor. The reported current costs incurred in the business year of the fund that ended on December 31st, 2011. They may change year by year. Current costs do not include performance fees and transaction costs. Further information about costs can be found in the prospectus in the chapter "costs".

## Past Performance



The past performance is not a reliable indicator for future performance. All charges and fees that have been incurred by Fund RED were subtracted in the calculation of the performance figures. Issue surcharges and redemption fees are not considered. Fund RED came into existence in 1993. The performance is calculated in EUR.

## Practical Information

The depository is ABC Bank Luxembourg S.A., Luxembourg. The prospectus, annual and semi-annual reports are available in the language of this document, as well as in English at XY Fund Company, S.A. and can be requested at no charge. You can obtain the latest unit prices and further information regarding this fund in the language of this document as well as in English on the local website of XY Fund Company, S.A. The fund is subject to Luxembourgian tax law. This may have an effect on how your personal income from the fund is taxed. XY Fund Company, S.A. can be held liable solely on the basis of any statement that is contained in this document which is misleading, inaccurate, or inconsistent with the relevant parts of the prospectus for the fund. This fund is authorized in Luxembourg and regulated by the Commission de Surveillance du Secteur Financier.

## Appendix 3.I / KID Fund WHITE

### Key Investor Information

This document provides you with key investor information about this fund. It is not marketing material. The information is required by law to help you understand the nature and the risks of investing in this fund. You are advised to read it so you can make an informed decision about whether to invest.

### Fund WHITE

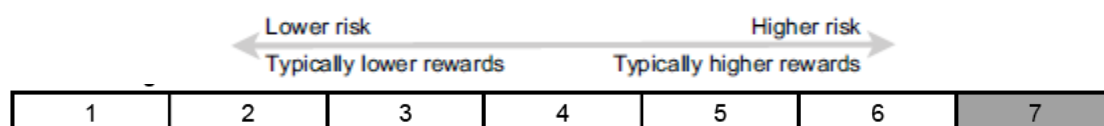
WKN: 555555 ISIN: LU000001110

Managed by XY Fund Company S.A.

### Objectives and Investment Policy

The objective of the investment policy is to sustainably generate additional capital gain compared to the benchmark (MSCI Europe (RI)). To achieve this, the fund mainly invests into stocks of established, highly capitalized European companies. Stocks of promising medium-sized and smaller companies with a long-term potential to advance are selectively added alongside. From the fund management's perspective, stocks of companies with business models that are undervalued by the market or underestimated growth perspectives are in the focus of the stock selection. Therefore, company specific criteria like for example a strong market position, promising products, a competent management, strategic focus on core competences, the return-oriented use of resources, a sustainably above-average profit development, as well as a shareholder-oriented information policy are in the focus when selecting positions. At the same time, valuation criteria in stock analysis influence investment decisions. Within this framework, the fund management is responsible for the selection of single investments. The fund is geared to a benchmark. It does not replicate the benchmark, but tries to outperform it and could therefore deviate considerably – positively and negatively – from the benchmark. The fund is subject to various risks. A detailed risk assessment and further remarks can be found in the prospectus in the chapter "risks". The currency of Fund WHITE is EUR. Income from the fund is paid out yearly. You can claim the redemption of fund units at each valuation date. Redemption may only be suspended in exceptional cases and in due consideration of your interest as an investor.

### Risk and Reward Profile



The calculation of the risk and reward profile is based on historical data that cannot be taken as a reliable basis for the future risk profile. This risk indicator is subject to changes; the classification of the fund may change over time and cannot be guaranteed. Even a fund that is classified the lowest category (category 1) is not a completely risk-free investment. The fund is classified as category 7, because unit prices can fluctuate very strongly, and chances of both, losses and gains, can be very high.

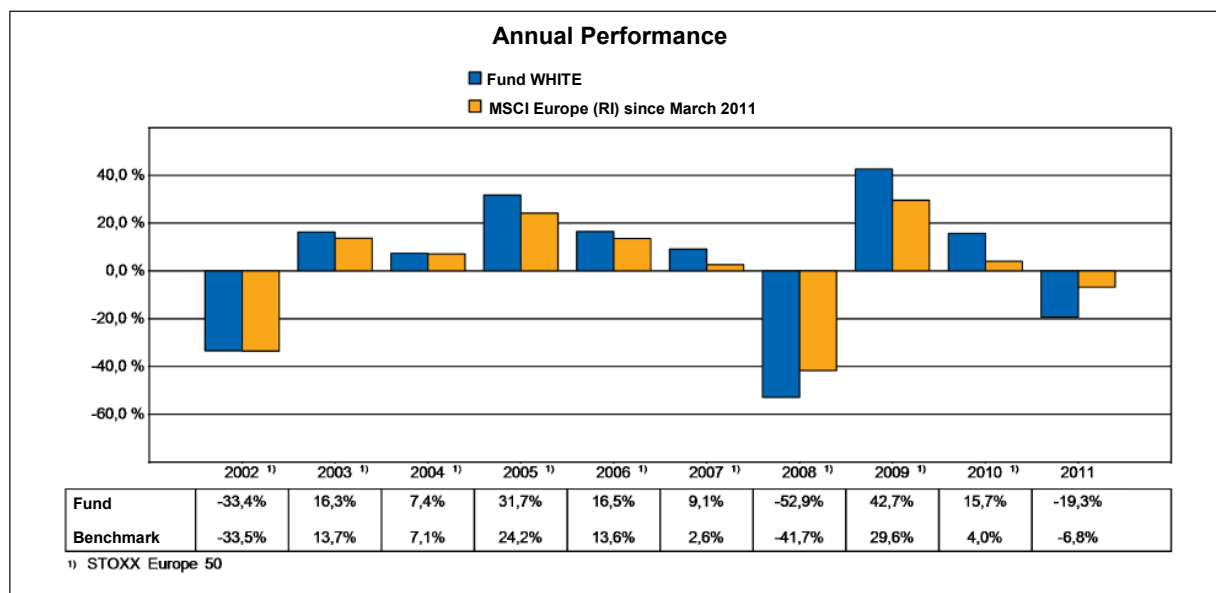
## Charges for this fund

The charges you pay are used to pay the costs of running the fund, including the costs of marketing and distributing it. These charges reduce the potential growth of your investment.

One-time costs pre and past investing	
Issue surcharges	5.00%. This is the maximum amount that is deducted from your investment prior to investing
Redemption fees	No redemption fees
Costs that are deducted annually from the fund	
Current Costs	1.40%
Costs the fund has to bear under certain circumstances	
Fees tied to the fund's performance	None
Compensation from securities lending	0.08%

The indicated issue surcharge is the maximum amount. It may be lower in a particular case. The amount in effect for you can be obtained from the department which is responsible for you or from your financial advisor. The reported current costs incurred in the business year of the fund that ended on September 31st, 2011. They may change year by year. Current costs do not include performance fees and transaction costs. Further information about costs can be found in the prospectus in the chapter "costs".

## Past Performance



The past performance is not a reliable indicator for future performance. All charges and fees that have been incurred by Fund WHITE were subtracted in the calculation of the performance figures. Issue surcharges and redemption fees are not considered. Fund WHITE came into existence in 1988. The performance is calculated in EUR.

## Practical Information

The depository is ABC Bank Luxembourg S.A., Luxembourg. The prospectus, annual and semi-annual reports are available in the language of this document, as well as in English at XY Fund Company, S.A. and can be requested at no charge. You can obtain the latest unit prices and further information regarding this fund in the language of this document as well as in English on the local website of XY Fund Company, S.A. The fund is subject to Luxembourgian tax law. This may have an effect on how your personal income from the fund is taxed. XY Fund Company, S.A. can be held liable solely on the basis of any statement that is contained in this document which is misleading, inaccurate, or inconsistent with the relevant parts of the prospectus for the fund. This fund is authorized in Luxembourg and regulated by the Commission de Surveillance du Secteur Financier.

## Appendix 3.J / Supplementary Descriptive Statistics Investment Stage

**Importance of Investment Policy**

	full sample	Treatment A	Treatment B	Treatment C	Treatment D
<b>very important (=1)</b>	46	10	18	9	9
<b>important (=2)</b>	59	14	9	19	17
<b>fairly important (=3)</b>	28	10	6	5	7
<b>less important (=4)</b>	18	11	3	1	3
<b>not important (=5)</b>	4	1	0	2	1
<b>don't know / no statement</b>	4	1	0	2	1
<b>total</b>	159	47	36	38	38
<b>Mean</b>	2.19	2.54	1.83	2.11	2.19
<b>Standard Error</b>	0.860	0.169	0.167	0.168	0.164
<b>Median</b>	2.00	2.00	1.50	2.00	2.00

Note: The statistics Mean, Standard Error, and Median do not include observations in the “don’t know / no statement” category.

**Importance of Performance**

	full sample	Treatment A	Treatment B	Treatment C	Treatment D
<b>very important (=1)</b>	37	12	7	9	9
<b>important (=2)</b>	60	20	15	15	10
<b>fairly important (=3)</b>	37	10	11	9	7
<b>less important (=4)</b>	13	2	2	2	7
<b>not important (=5)</b>	8	3	0	2	3
<b>don't know / no statement</b>	4	0	1	1	2
<b>total</b>	159	47	36	38	38
<b>Mean</b>	2.32	2.23	2.23	2.27	2.58
<b>Standard Error</b>	0.087	0.159	0.143	0.176	0.216
<b>Median</b>	2.00	2.00	2.00	2.00	2.00

Note: The statistics Mean, Standard Error, and Median do not include observations in the “don’t know / no statement” category.

#### Importance of Volatility

	full sample	Treatment A	Treatment B	Treatment C	Treatment D
<b>very important (=1)</b>	27	8	7	8	4
<b>important (=2)</b>	55	18	13	12	12
<b>fairly important (=3)</b>	44	12	10	13	9
<b>less important (=4)</b>	21	6	5	2	8
<b>not important (=5)</b>	9	3	1	2	3
<b>don't know / no statement</b>	3	0	0	1	2
<b>total</b>	159	47	36	38	38
<b>Mean</b>	2.55	2.53	2.44	2.41	2.83
<b>Standard Error</b>	0.088	0.163	0.176	0.175	0.193
<b>Median</b>	2.00	2.00	2.00	2.00	3.00

Note: The statistics Mean, Standard Error, and Median do not include observations in the “don’t know / no statement” category.

#### Importance of Asset Preservation

	full sample	Treatment A	Treatment B	Treatment C	Treatment D
<b>very important (=1)</b>	46	10	14	10	12
<b>important (=2)</b>	64	23	14	15	12
<b>fairly important (=3)</b>	27	9	6	6	6
<b>less important (=4)</b>	14	4	1	4	5
<b>not important (=5)</b>	4	1	1	1	1
<b>don't know / no statement</b>	4	0	0	2	2
<b>total</b>	159	47	36	38	38
<b>Mean</b>	2.14	2.21	1.92	2.19	2.19
<b>Standard Error</b>	0.082	0.139	0.161	0.177	0.190
<b>Median</b>	2.00	2.00	2.00	2.00	2.00

Note: The statistics Mean, Standard Error, and Median do not include observations in the “don’t know / no statement” category.



**Importance of Verbal Description of Risk and Reward Profile**

	full sample	Treatment A	Treatment B	Treatment C	Treatment D
<b>very important (=1)</b>	9	3	6		
<b>important (=2)</b>	22	12	10		
<b>fairly important (=3)</b>	21	13	8		
<b>less important (=4)</b>	20	12	8	n/a	n/a
<b>not important (=5)</b>	9	6	3		
<b>don't know / no statement</b>	2	1	1		
<b>total</b>	83	47	36		
<b>Mean</b>	2.98	3.13	2.77		
<b>Standard Error</b>	0.133	0.169	0.209	n/a	n/a
<b>Median</b>	3.00	3.00	3.00		

Note: The statistics Mean, Standard Error, and Median do not include observations in the “don’t know / no statement” category.

**Importance of Costs**

	full sample	Treatment A	Treatment B	Treatment C	Treatment D
<b>very important (=1)</b>	33	11	6	7	9
<b>important (=2)</b>	43	12	12	10	9
<b>fairly important (=3)</b>	45	15	9	10	11
<b>less important (=4)</b>	21	5	6	6	4
<b>not important (=5)</b>	14	4	3	3	4
<b>don't know / no statement</b>	3	0	0	2	1
<b>total</b>	159	47	36	38	38
<b>Mean</b>	2.62	2.55	2.67	2.67	2.59
<b>Standard Error</b>	0.097	0.177	0.199	0.203	0.210
<b>Median</b>	3.00	3.00	2.50	3.00	3.00

Note: The statistics Mean, Standard Error, and Median do not include observations in the “don’t know / no statement” category.

#### Importance of SRRI Values

	full sample	Treatment A	Treatment B	Treatment C	Treatment D
<b>very important (=1)</b>	19	10	n/a	9	n/a
<b>important (=2)</b>	35	18		17	
<b>fairly important (=3)</b>	17	11		6	
<b>less important (=4)</b>	7	3		4	
<b>not important (=5)</b>	5	4		1	
<b>don't know / no statement</b>	2	1		1	
<b>total</b>	85	47		38	
<b>Mean</b>	2.33	2.41		2.22	
<b>Standard Error</b>	0.121	0.172		0.170	
<b>Median</b>	2.00	2.00		2.00	

Note: The statistics Mean, Standard Error, and Median do not include observations in the “don’t know / no statement” category.

#### Importance of Portfolio Structure

	full sample	Treatment A	Treatment B	Treatment C	Treatment D
<b>very important (=1)</b>	32	8	13	5	6
<b>important (=2)</b>	57	19	13	12	13
<b>fairly important (=3)</b>	33	11	7	8	7
<b>less important (=4)</b>	20	4	2	8	6
<b>not important (=5)</b>	8	3	1	2	2
<b>don't know / no statement</b>	9	2	0	3	4
<b>total</b>	159	47	36	38	38
<b>Mean</b>	2.43	2.44	2.03	2.71	2.56
<b>Standard Error</b>	0.092	0.164	0.171	0.195	0.199
<b>Median</b>	2.00	2.00	2.00	3.00	2.00

Note: The statistics Mean, Standard Error, and Median do not include observations in the “don’t know / no statement” category.

## Appendix 3.K / Supplementary Descriptive Statistics KID Assessment Stage

**I was able to compare the objectives and investment policies of the funds without any problems**

	full sample	Treatment A	Treatment B	Treatment C	Treatment D
not agree at all (=1)	12	3	1	2	6
partly not agree (=2)	39	13	10	7	9
undecided (=3)	27	9	1	10	7
partly agree (=4)	63	19	22	11	11
fully agree (=5)	18	3	2	8	5
<b>total</b>	159	47	36	38	38
<b>Mean</b>	3.23	3.13	3.39	3.42	3.00
<b>Standard Error</b>	0.092	0.160	0.175	0.191	0.213
<b>Median</b>	4.00	3.00	4.00	3.50	3.00

**It was particularly important for my investment decision to compare the objectives and investment policies of the funds**

	full sample	Treatment A	Treatment B	Treatment C	Treatment D
not agree at all (=1)	7	2	0	3	2
partly not agree (=2)	24	7	3	9	5
undecided (=3)	25	12	3	5	5
partly agree (=4)	55	17	15	8	15
fully agree (=5)	48	9	15	13	11
<b>total</b>	159	47	36	38	38
<b>Mean</b>	3.71	3.51	4.17	3.50	3.74
<b>Standard Error</b>	0.093	0.161	0.152	0.226	0.191
<b>Median</b>	4.00	4.00	4.00	4.00	4.00

**I was able to compare the risk and reward profiles of the funds without any problems**

	full sample	Treatment A	Treatment B	Treatment C	Treatment D
not agree at all (=1)	15	4	3	2	6
partly not agree (=2)	33	5	7	12	9
undecided (=3)	19	4	7	6	2
partly agree (=4)	59	21	14	8	16
fully agree (=5)	33	13	5	10	5
total	159	47	36	38	38
Mean	3.39	3.72	3.31	3.32	3.13
Standard Error	0.102	0.179	0.198	0.214	0.220
Median	4.00	4.00	4.00	3.00	4.00

**It was particularly important for my investment decision to compare the risk and reward profiles of the funds**

	full sample	Treatment A	Treatment B	Treatment C	Treatment D
not agree at all (=1)	1	0	0	0	1
partly not agree (=2)	10	1	2	3	4
undecided (=3)	22	8	6	4	4
partly agree (=4)	56	19	11	9	17
fully agree (=5)	70	19	17	22	12
total	159	47	36	38	38
Mean	4.16	4.19	4.19	4.32	3.92
Standard Error	0.074	0.116	0.153	0.156	0.170
Median	4.00	4.00	4.00	5.00	4.00

**I was able to compare the past financial performances of the funds without any problems**

	full sample	Treatment A	Treatment B	Treatment C	Treatment D
not agree at all (=1)	9	3	3	1	2
partly not agree (=2)	17	4	2	7	4
undecided (=3)	22	10	6	4	2
partly agree (=4)	55	16	11	12	16
fully agree (=5)	56	14	14	14	14
total	159	47	36	38	38
Mean	3.83	3.72	3.86	3.82	3.95
Standard Error	0.094	0.171	0.208	0.195	0.188
Median	4.00	4.00	4.00	4.00	4.00

**It was particularly important for my investment decision to compare the past financial performances of the funds**

	full sample	Treatment A	Treatment B	Treatment C	Treatment D
not agree at all (=1)	8	3	2	1	2
partly not agree (=2)	22	5	4	4	9
undecided (=3)	24	9	5	8	2
partly agree (=4)	55	14	15	11	15
fully agree (=5)	50	16	10	14	10
total	159	47	36	38	38
Mean	3.74	3.74	3.75	3.87	3.58
Standard Error	0.094	0.179	0.193	0.182	0.205
Median	4.00	4.00	4.00	4.00	4.00

**I was able to compare the costs of the funds without any problems**

	full sample	Treatment A	Treatment B	Treatment C	Treatment D
not agree at all (=1)	7	2	1	2	2
partly not agree (=2)	24	5	7	10	2
undecided (=3)	22	6	5	6	5
partly agree (=4)	51	16	13	11	11
fully agree (=5)	55	18	10	9	18
total	159	47	36	38	38
Mean	3.77	3.91	3.67	3.39	4.08
Standard Error	0.095	0.169	0.195	0.205	0.186
Median	4.00	4.00	4.00	4.00	4.00

**It was particularly important for my investment decision to compare the costs of the funds**

	full sample	Treatment A	Treatment B	Treatment C	Treatment D
not agree at all (=1)	10	0	4	3	3
partly not agree (=2)	27	8	1	8	10
undecided (=3)	35	13	10	4	8
partly agree (=4)	56	17	14	14	11
fully agree (=5)	31	9	7	9	6
total	159	47	36	38	38
Mean	3.45	3.57	3.53	3.47	3.18
Standard Error	0.093	0.145	0.197	0.209	0.199
Median	4.00	4.00	4.00	4.00	3.00

**I was able to compare the funds in general without any problems**

	<b>full sample</b>	<b>Treatment A</b>	<b>Treatment B</b>	<b>Treatment C</b>	<b>Treatment D</b>
<b>not agree at all (=1)</b>	13	4	2	2	5
<b>partly not agree (=2)</b>	32	8	6	9	9
<b>undecided (=3)</b>	31	10	5	8	8
<b>partly agree (=4)</b>	66	20	19	14	13
<b>fully agree (=5)</b>	17	5	4	5	3
<b>total</b>	159	47	36	38	38
<b>Mean</b>	3.26	3.30	3.47	3.29	3.00
<b>Standard Error</b>	0.091	0.166	0.180	0.184	0.196
<b>Median</b>	4.00	4.00	4.00	3.50	3.00

**It was particularly important for my investment decision to compare the funds in general**

	<b>full sample</b>	<b>Treatment A</b>	<b>Treatment B</b>	<b>Treatment C</b>	<b>Treatment D</b>
<b>not agree at all (=1)</b>	1	0	0	0	1
<b>partly not agree (=2)</b>	7	2	1	3	1
<b>undecided (=3)</b>	17	3	5	4	5
<b>partly agree (=4)</b>	72	30	12	12	18
<b>fully agree (=5)</b>	62	12	18	19	13
<b>total</b>	159	47	36	38	38
<b>Mean</b>	4.18	4.11	4.31	4.24	4.08
<b>Standard Error</b>	0.066	0.102	0.137	0.153	0.148
<b>Median</b>	4.00	4.00	4.50	4.50	4.00

**Actually I would have needed to contact an investment advisor prior to my decision**

	<b>full sample</b>	<b>Treatment A</b>	<b>Treatment B</b>	<b>Treatment C</b>	<b>Treatment D</b>
<b>not agree at all (=1)</b>	29	11	8	5	5
<b>partly not agree (=2)</b>	16	3	4	6	3
<b>undecided (=3)</b>	15	5	3	2	5
<b>partly agree (=4)</b>	30	11	7	6	6
<b>fully agree (=5)</b>	69	17	14	19	19
<b>total</b>	159	47	36	38	38
<b>Mean</b>	3.59	3.43	3.42	3.74	3.82
<b>Standard Error</b>	0.123	0.233	0.271	0.249	0.238
<b>Median</b>	4.00	4.00	4.00	4.50	4.50

**I would have liked to acquire further information about the available funds prior to my decision**

	<b>full sample</b>	<b>Treatment A</b>	<b>Treatment B</b>	<b>Treatment C</b>	<b>Treatment D</b>
<b>not agree at all (=1)</b>	11	5	1	4	1
<b>partly not agree (=2)</b>	11	3	3	3	2
<b>undecided (=3)</b>	10	3	2	2	3
<b>partly agree (=4)</b>	49	13	12	9	15
<b>fully agree (=5)</b>	78	23	18	20	17
<b>total</b>	159	47	36	38	38
<b>Mean</b>	4.08	3.98	4.19	4.00	4.18
<b>Standard Error</b>	0.096	0.196	0.177	0.223	0.159
<b>Median</b>	4.00	4.00	4.50	5.00	4.00

## 4 Making Easy Money – Do We Trade Risk for Laziness?<sup>59</sup>

### 4.1 Introduction

In many industries, variable components represent a large share of total wages. By linking “success” – however it is defined in a specific case – to higher salary, a variable payment scheme is intended to incentivize employees to work hard to reach their targets, i.e. to exert high effort to fulfill their designated tasks. The simplest form of variable compensation is a piece rate that increases a worker’s salary by a constant amount for each unit that she produces. Thus, the total payment is a linear function of the worker’s output, and the more effort she exerts, the more she earns. Another variable payment scheme that is often used in executive pay uses bonus payments. In contrast to simple piece rates, bonus payments are made contingent on certain threshold values, target figures, or events. In the simplest form of a bonus payment, a manager earns the bonus if she has reached a defined target, for example if her business unit has contributed at least the annual profit that is specified in the manager’s contract. She does not earn the bonus if she falls short of reaching the target. Another common form of bonus systems is to install tournaments. For example, the best ten sales agents receive a bonus payment, while all the others do not. Instead of defining just one target value to determine if a manager is eligible for a bonus payment, it is also possible to specify several intervals of the target value that qualify for increasing amounts of bonus payments. The crucial distinction between an ordinary piece rate and a bonus system is that under a bonus system, the variable earnings of a manager increase disproportionately high with higher output levels, or, in general, degrees of target achievement. Thus, the total payment is a convex function of the manager’s output.

Bonus systems with such convex payment structures are very common in the financial industry, and in the context of the ongoing academic, political, and public dispute about the global banking crisis they have been heavily criticized for inducing undesired risk-taking: managers who fear that they might not earn their highly attractive bonus payment at the end of the year might be tempted to increase the risk of their business strategy, e.g. by leveraging

---

<sup>59</sup> The original essay, entitled “Making Easy Money – Do We Trade Risk for Laziness?”, is joint research work of Stefan Trautmann, Gijs van de Kuilen, and me. See Trautmann et al. (2013), unpublished manuscript.



their portfolios or reducing diversification and hope for the best. Even worse, they might have never intended to actually work hard to create sustainable value for their investors or companies, but all they ever wanted was making easy money through gambling instead of decent work.<sup>60</sup>

In this article, we look at these two competing motives for risk-taking using an experimental approach. Is excessive risk-taking driven by the greed to get rich quickly and effortlessly, therefore being the spawn of the agents' extreme preferences, or is it merely a consequence of extremely skewed incentives that simply make it attractive to take on risk even if one is not particularly greedy? We answer this question by comparing the effort provision levels in a laboratory experiment with two conditions: one offers the opportunity to choose a risky outside option, the other one does not. We find that risk-taking is indeed very common whenever the outside option is available, but the vast majority of the subjects who choose to gamble do not exert significantly less effort. Thus, they do not make an *ex ante* choice to cut down effort and to make easy money by only taking on risk. We do not analyze whether bonus systems actually succeed to incentivize higher effort than, e.g. ordinary piece rates or fixed payments. Our result, however, indicates that both, the opportunity to take on risk and – except for the few cases of very excessive usage of the outside option – even actual risk-taking activity do not reduce effort provision. Eriksson and Villeval (2008) and Dohmen and Falk (2011) show experimentally that effort provision increases under a variable payment scheme compared to fixed payments. Neither of these studies applies convex payment schemes, which should *per se* offer even higher incentives to exert high effort. Van Dijk et al. (2001) compare effort levels for individual, team, and tournament incentives and find that effort is higher under tournaments than in the other two conditions. Note that the payoff function in a tournament is also convex for each individual, as higher effort increases the probability of winning the tournament and to earn the bonus. However, this means that there is a payoff uncertainty inherent to tournaments, which is not the case in our bonus system, because individual payments do not depend on other agents' actions. For our experiment, we use the same task as Gill et al. (2013), who find that a randomly assigned bonus in the first round does not increase productivity in the second round. Instead, subjects in the bonus treatment are found more likely to cheat on the experimenter, even if they did not actually

---

<sup>60</sup> Ross (2004) points out that not all types of agents will become more risk seeking or less risk averse under a convex payment scheme. He shows theoretically, that while convexity always makes risk more desirable, behavior also depends on the shape of the individual utility function. With increasing risk aversion, the convexity effect may be offset.

receive the bonus in the first round. Cheating, in their experiment, has nothing to do with risk-taking. The subjects report presumably false, non-verifiable phone numbers in order to increase their payment.

Our result suggests that risk-taking is not deliberately planned *ex ante*, but it results from an *ex post* trade-off between the, by then, certain amount that has been earned under the bonus system and the potential gain from the risky outside option. Such a decision should be governed by an individual's risk preference and the value constellation of the lottery's certainty equivalent and the payment from the bonus system.<sup>61</sup> Furthermore, this choice should not depend on the payment system that generates the certain payoff. Whether the certain payoff is earned under a convex or a fixed (i.e. not dependent on effort provision) payment scheme should not make a difference: as long as the risky outside option is more attractive, it should be chosen. This implies that measures to reduce or effectively prohibit opportunities to take on risk should be given priority to banning bonus systems when designing governance structures to control managerial behavior in the aspect of risk-taking.

Interestingly, we find that the *ex post* choices are often even too cautious in the light of the previously elicited risk preferences. The convex payment structure has the effect that the (*ex post*) certain payment is only attractive when a high effort level is achieved. However, a significant number of test persons do not choose the risky outside option even if the elicited certainty equivalent is higher. This could be interpreted as an additional sign that greed in the sense of the desire to make easy money is not a prevailing motive for risk-taking.

The next section describes our experimental setup in detail. Section 4.3 presents the results and Section 4.4 concludes with a discussion and suggestions for further research.

## **4.2 Experimental Design**

We apply a between-subjects design with a treatment group and a control group. In the treatment group, the subjects can make *ex post* choices between two different payment modes in some stages. These choices are not available in the control group. First, we will describe the general setup of the experiment (i.e. the control group design) and afterwards the treatment variation.

---

<sup>61</sup> In an *ex post* decision, the costs of effort provision are sunk costs and should not be taken into account anymore. In an *ex ante* decision, a subject should consider that effort is costly, as it is quite troublesome to achieve high scores in the task in very limited time.

The experiment was programmed and conducted using the software z-Tree (Fischbacher, 2007). The sessions were conducted in January 2013 in the CentER lab at Tilburg University. Subjects were recruited via the email list of CentER lab. Each participant received a show up fee of seven Euros in addition to what she earned in the experiment. In total, we had 70 subjects (37 in the treatment group and 33 in the control group). Before the start of each session, all participants were given enough time to read a general instruction handout about the structure and content of the experiment (see Appendix 4.A).

In the first stage, we determine each subject's certainty equivalent  $CE_i$  for the binary lottery  $L$ . The lottery pays EUR 12.50 with a probability of 30 percent and zero otherwise. The subjects are presented two choice lists where they decide between playing the lottery and receiving increasing amounts of money. In the first list, the sure payoffs increase in steps of 50 cents, starting at zero. We enforce that subjects make consistent decisions by switching from the lottery to the sure payoff only once. In the second list, we zoom into the interval in which a subject switched from playing the lottery to accepting the sure payoff in the first list and the procedure is repeated with gradual increases of two cents per step, allowing us to determine the exact monetary amount of  $CE_i$ . The test persons are instructed that one of the choices from one of the lists will be selected randomly and that they can earn the sure payoff or the outcome of the lottery, depending on their actual choice. Appendix 4.B and Appendix 4.C contain the instructions and screenshots of the choice lists.

In stages 2 – 6 the subjects work on the “slider task”, a real-effort task developed by Gill and Prose (2011). In a limited timeframe, the subjects have to move sliders to the central position by using the computer mouse in order to earn money. Before the start of stage 2, the participants are instructed about the task and the payment mode for this round, followed by four screenshots to familiarize with the structure of the subsequent screen without losing time during the actual task. Appendices 4.D through 4.H show the detailed instructions and the screenshots. In the actual task, there are 48 sliders displayed in each round. Let  $e_{i,t}$  be the effort level of subject  $i$  in stage  $t$ , defined as the number of sliders that are correctly positioned in the middle within the allotted time. In each stage, subjects can earn  $e_{i,t}^2$  cents. Each participant has a printed list of effort levels and corresponding earnings on their desk throughout the whole experiment (see Appendix 4.I). This payment scheme is chosen to resemble a typical bonus system with regards to its convexity: it provides a strong incentive to reach high effort levels, while low effort levels are less remunerative. In stage 2, the subjects have 105 seconds to work on the slider task. In each subsequent stage, the allotted time is

increased by 8 seconds, resulting in 137 seconds in the final round. The participants are fully instructed about the time structure of the entire remaining experiment in the beginning of stage 2. We want to make sure that the test persons anticipate higher payoffs in later periods even if they forget to take training effects into account.<sup>62</sup> After each round of the slider task, the participants are informed about their effort level and the corresponding payment. In each stage, there is a “Skip” button available that terminates the slider task immediately. When skipping the current stage, a screen with amusing cartoons is displayed instead. The payment is determined by the effort level that has been achieved before leaving the task. We included the “Skip” option so that the test persons do not have to work on the slider task out of mere boredom if they actually do not want to exert effort, just because there is no alternative costless activity available. Before proceeding to the next stage, there is a 45 seconds intermission in which cartoons (different from the ones after skipping) are displayed on the screens so that the test persons can relax for a while. From stage 3 on (i.e. the second round of the slider task), the subjects are also asked to report their belief about the other participants’ average effort level in the current round before starting to work on the task. They can earn a reward of 5 Euros if their belief lies within a range of +1 / -1 around the actual value (see Appendix 4.J for a screenshot of the Belief Elicitation screen).

After the completion of the final stage, the total profits are calculated by randomly selecting one out of all six stages to determine the payment from either the slider task (stages 2 – 6) or the choice list (stage 1). Furthermore, one out of the last four stages (stages 3 – 6) is randomly selected to determine the payment from the belief elicitation task. The subjects are informed about their remuneration from the two selected stages, and their total earnings including the show up fee. We paid the participants on the same day via wire transfer. The experiment concludes with a short questionnaire in which the test persons enter their bank details, their age, nationality, and whether they are a student at the Tilburg School of Economics and Management or a different school.

In the experimental treatment group, the subjects can choose whether they would like to receive the payment of  $e_{i,t}^2$  cents or play lottery  $L$  instead. This choice is first available in stage 3 (i.e. the second round of the slider task) and subjects are informed about the option before they start working on the task (see Appendix 4.K for the complete instructions). Furthermore, they are informed on a separate screen that “in some, but not necessarily all of

---

<sup>62</sup> The learning potential is substantial in the slider task. The average number of sliders that is positioned correctly per second almost doubles during the course of the experiment from about 0.08 in the first round to 0.16 in the last round.

the following rounds, you can also choose to play the gamble instead of receiving your earnings from these rounds. If the gambling option is available in later rounds, the gamble will always be exactly the same as the one in this round.” In fact, the lottery option is available in stages 3, 4, and 5 for all subjects in the treatment group. Thus, the first (stage 2) and the last round (stage 6) of the slider task are identical for both groups, while the three rounds in between offer the lottery option for the treatment group only. The subjects make the choices between the lottery and the sure payoff that they can earn after the completion of the slider task in each round (see Appendix 4.L). The first round of the slider task (stage 2) serves as a measure of each individual’s unbiased base productivity, as both experimental conditions are identical up to this point. The last round (stage 6) is intended to measure effort levels under equal circumstances that are preceded by different choice histories. Thus, we can detect potential long-term effects on productivity caused by the availability of the lottery option.

We designed the lottery  $L$  in a way that it is attractive, but not normatively appealing to all subjects. On the one hand, the payoff of 1,250 cents is very high compared to the amounts that can be earned in the slider task. It corresponds to an effort level between 35 and 36, which is hardly achievable in the slider task.<sup>63</sup> On the other hand, the lottery is heavily skewed in favor of the zero-outcome. Therefore, only for subjects with very low effort levels the lottery constitutes the clearly better alternative.

## 4.3 Results

### 4.3.1 Description of the Samples

Table 4.1 provides an overview of the demographics of the two samples. The number of subjects in the treatment group is slightly larger than in the control group. The subjects in the control group are on average a little bit more risk seeking than in the treatment group (the expected value of lottery  $L$  is only 375 cents). However, when transforming the respective certainty equivalents into the effort level that is necessary to earn the respective amounts under the applied payment scheme in the slider task, they both fall into the same interval between 22 and 23 sliders. Furthermore, a Mann-Whitney U-Test does not indicate a significant difference between the two means. The portions of male and Dutch subjects and of

---

<sup>63</sup> Among all 350 observations (70 test persons over five rounds), there are only four with an effort level equal to or above 30 (1x 30, 1x32, 1x35, 1x37).

those from the school of economics are higher in the treatment group. Among these, we find that gender and subject of study have a significant influence on behavior in the experiment.

Table 4.1: Demographics

	<b>Treatment Group</b>	<b>Control Group</b>	<b>Total</b>
<b>N</b>	37	33	70
<b>CertaintyEquivalent</b>	484,95	512,58	497,97
<b>Age</b>	22,73	22,67	22,7
<b># Male Subjects</b>	21 (56.8%)	15 (45.5%)	36
<b># Female Subjects</b>	16 (43.2%)	18 (54.5%)	34
<b># Dutch Students</b>	18 (48.5%)	11 (33.3%)	29
<b># Other Nationalities</b>	19 (51.5%)	22 (66.7%)	41
<b># Econstudents</b>	30 (81.1%)	22 (66.7%)	52
<b># Other Fields of Study</b>	7 (18.9%)	11 (33.3%)	18

#### 4.3.2 Effort Exertion and Lottery Choices

Table 4.2 shows the effort levels in the five stages for the treatment and the control group. Even though we observe that the average effort levels are always a little bit higher in the control group, the differences are not statistically significant at the 90 percent confidence level, as indicated by pairwise independent t-tests for stages 2, 3, 4, and 6 and a Mann-Whitney U-Test for stage 5.

Table 4.2: Effort Levels by Stage and Experimental Group

	Treatment Group		Control Group	
	Mean	Standard Error	Mean	Standard Error
<b>Effort stage 2</b>	8.78 <sub>a</sub>	0.83	9.18 <sub>a</sub>	0.78
<b>Effort stage 3</b>	13.54 <sub>a</sub>	0.79	14.45 <sub>a</sub>	0.86
<b>Effort stage 4</b>	15.70 <sub>a</sub>	1.06	17.33 <sub>a</sub>	0.99
<b>Effort stage 5</b>	17.97 <sub>a</sub>	1.18	19.15 <sub>a</sub>	1.02
<b>Effort stage 6</b>	21.95 <sub>a</sub>	0.71	21.88 <sub>a</sub>	0.86

Note: Values in the same row *not* sharing the same subscript are significantly different at  $p < 0.1$  in the test of equality for column means. For stages 2, 3, 4, and 6 we use independent t-tests. For stage 5, we use a non-parametric Mann-Whitney U-Test, because the data is not normally distributed in any of the two groups, as indicated by Kolmogorov-Smirnov-Tests for normality (Treatment Group:  $Z(37) = 1.396$ ,  $p = 0.041$ , Control Group:  $Z(33) = 1.379$ ,  $p = 0.045$ ).

Interestingly, this is the case even though the lottery option is frequently chosen among the subjects in the treatment group. In stage 3, when the lottery option is available for the first time, 23 subjects (62.2%) prefer to play the lottery over receiving the sure payoff that they had earned in the slider task before. In stage 4, the number of lottery choices decreases to 18 (48.7% of all subjects), and in the last stage in which the lottery option is available, 11 subjects (29.7%) still choose to gamble, while on average the effort level increases considerably over time. A pooled regression of the effort levels in stages 3, 4, and 5 (when the lottery option is available) on a comprehensive set of variables confirms that subjects in the treatment group do not behave significantly different from subjects in the control group. Table 4.3 summarizes the results.

Table 4.3: Pooled Regression Effort Levels Stages 3, 4, 5

Effort (stage 3, 4, 5)	Coefficient	Robust SE	t-value	p-value
<b>Stage3</b>	11.170**	4.857	2.30	0.025
<b>Stage4</b>	13.670***	4.997	2.74	0.008
<b>Stage5</b>	15.727***	4.966	3.17	0.002
<b>Treatment</b>	-1.374	1.065	-1.29	0.201
<b>CertaintyEquivalent</b>	0.002	0.002	0.67	0.505
<b>Effort_firstround</b>	0.572***	0.149	3.84	0.000
<b>Age</b>	-0.182	0.201	-0.91	0.369
<b>Male</b>	1.971*	1.106	1.78	0.079
<b>Dutch</b>	0.573	1.248	0.46	0.648
<b>Econstudent</b>	0.714	1.456	0.49	0.625

Note: Number of observations = 210. Standard errors are clustered on subject level to account for repeated observations. Standard errors are robust against heteroskedasticity. \*\*\*, \*\*, \* indicate statistical significance at the 99%, 95%, and 90% confidence levels.  $F(10, 69) = 203.60$ ,  $p = 0.000$ .  $R^2 = 0.9190$ .

The dummy variables “*Stage3*”, “*Stage4*”, and “*Stage5*” capture differences between the three stages that stem from additional time, training effects or fatigue. We see that effort levels generally increase over periods. The regression coefficient of the “*Treatment*” dummy (0 = control group without lottery option; 1 = treatment group with lottery option) is negative, but not significantly different from 0 at the 90 percent confidence level. Thus, the regression confirms that the availability of the lottery option does not have a significant influence on the effort that is exerted by the subjects. Individual differences in risk preferences, as indicated by “*CertaintyEquivalent*”, also do not have any influence on effort levels. The effort level of the first round of the slider task (stage 2, “*Effort\_firstround*”) has a significant positive influence on subsequent performance.<sup>64</sup> This indicates that there are persistent individual differences between subjects that are probably due to unobserved factors like motivation, eye-sight, or skillfulness in using a computer mouse.<sup>65</sup> Male subjects exert significantly more effort than females (about two sliders more per round). “*Age*” (in years), nationality (binary coded,

<sup>64</sup> Replacing the variable “*Effort\_fistround*” by “*Effort\_previousround*”, i.e. using the effort level from the stage immediately before the current stage yields qualitatively the same results.

<sup>65</sup> A regression with interaction terms for the stage dummies with the variables “*Treatment*”, “*CertaintyEquivalent*”, “*Effort\_firstround*” and “*Effort\_previousround*” yields qualitatively the same results. None of the interaction terms is significantly different from zero, i.e. there are no effects of these variables on effort exertion that are stage-specific. Therefore we only report the main effects. Additional tables are available upon request.



“*Dutch*” = 1 vs. “*Non-Dutch*” = 0) and the field of study (binary coded, “*Econstudent*” = 1 vs. “*Econstudent*” = 0) do not significantly influence effort.

Remember that the subjects know whether the lottery option is available in the current stage before they start working on the slider task, but the actual choice is made ex post. The fact that the effort levels are not significantly different between both treatments reveals that on average the subjects in the experimental group do not anticipate their later choice behavior, and thus there is no detrimental ex ante effect on effort. If a subject knows that she will choose to play the lottery at the end of the stage, she should not exert costly effort, but rather skip the task and look at the cartoons instead.<sup>66</sup> However, the “Skip” option is only used five times in the treatment group (twice in stage 4 and three times in stage 5). No subject in the control group decided to ever skip the slider task early.

However, even though there is no general treatment effect on effort exertion on average, there are of course behavioral differences with regards to effort and lottery choices within the experimental group, i.e. among the test persons that have the lottery option at hand, that are worth to examine further.

Our results suggest that lottery choices are made ex post by the subjects, because they do not cut down their effort significantly if the lottery is available and they almost never skip the task. Technically spoken, the monetary gain from the effort task could be treated as an exogenous regressor to explain choice behavior with regards to the lottery: after the completion of the task, a test person compares the payment that she can earn with her certainty equivalent for the lottery and makes her choice. The payment from the task is given at the moment when the choice is made and it therefore qualifies as an independent variable in a regression model. However, this is a mere conjecture, and we need to test it by assuming the opposite and treating the payment from the effort task as an endogenous regressor and see if our hypothesis of exogeneity holds or needs to be rejected.

We run a probit regression of the dummy variable “*Choselottery*” (0 = subject chooses payment earned in the slider task; 1 = subject chooses to play the lottery) on the variables “*Paymenteffort*” (monetary amount in cents that can be earned according to individual effort level in the current round), “*CertaintyEquivalent*”, “*Age*”, “*Male*”, “*Dutch*”, and

---

<sup>66</sup> We cannot measure the exact cost or disutility of effort exertion, nor the utility of watching the cartoons. While both are likely to vary between persons, depending on personal skills, taste in cartoons and maybe other factors, we argue that it is reasonable to assume that watching the cartoon should always be preferred to working on the slider task, especially if the subject has already decided to play the lottery and therefore the enjoyment of earning money by exerting effort does not play a role.

“*Econstudent*”. Since we suspect “*Paymenteffort*” to be endogenous, i.e. it could be influenced by the dependent dummy variable in a way that a subject has already decided whether she would like to play the lottery or not before starting to work on the task, we use the payment from the task that each subject has earned in the previous round (“*Paymenteffort\_previousround*”) as an instrument in a recursive model that replaces the potentially endogenous variable by a linear combination of the instrument and the exogenous variables. The model parameters are obtained iteratively through maximum likelihood estimation. “*Paymenteffort\_previousround*” is a suitable instrument, because we know that effort levels are significantly influenced by the results of the previous rounds in all stages and it is certainly independent, since the test persons only learn about the availability of the lottery in the current round after the previous round is completed. Table 4.4 summarizes the results.

Table 4.4: Probit Regression Lottery Choice with endogenous regressor

Choselottery	Coefficient	Average marginal effect	Robust SE	z	P >  z
Constant	1.0345	-	2.0339	0.51	0.611
Paymenteffort	-0.0077***	-0.0019***	0.0019	-4.03	0.000
CertaintyEquivalent	-0.0017*	-0.0004**	0.0009	-1.92	0.055
Age	0.0332	0.0081	0.0777	0.43	0.669
Male	0.2780	0.0673	0.3010	0.92	0.356
Dutch	0.0065	0.0016	0.5039	0.01	0.990
Econstudent	1.1440**	0.2771***	0.4501	2.54	0.011

Note: Number of observations = 111 (Treatment group only, stages 3-5). Standard errors are clustered on subject level to account for repeated observations. Standard errors are robust against heteroskedasticity. \*\*\*, \*\*, \* indicate statistical significance at the 99%, 95%, and 90% confidence levels. “*Paymenteffort*” has been instrumented by “*Paymenteffort\_previousround*”, Wald test of exogeneity  $\chi^2(1) = 6.70$ ,  $p = 0.0096$ , Wald test of joint significance  $\chi^2(6) = 19.02$ ,  $p = 0.0041$ .

A Wald test of exogeneity does not support the hypothesis that “*Paymenteffort*” is exogenous. Exogeneity is rejected with a confidence level of more than 99 percent ( $\chi^2(1) = 6.70$ ,  $p = 0.0096$ ). Thus, even though we do not observe significantly lower effort levels (and correspondingly lower payments earned in the task) in the experimental group on average, we would go too far by concluding that the availability of the lottery option does not influence

effort exertion at all. We will discuss this in more detail further below – first, we turn to the other results presented in Table 4.4.

In addition to the obtained regression coefficients, we also report average marginal effects of changes in the independent variables on the dependent dummy variable. They are different from the coefficients, because a change in any independent variable does not only have an immediate effect, but also an indirect effect through the instrumented regressor. We observe significantly negative marginal effects for “*Paymenteffort*” and “*CertaintyEquivalent*”, which indicates that the probability of the lottery being chosen decreases with higher values of these two control variables. Notably, economics students exhibit a 27 percent higher probability of choosing the lottery than students from other faculties. Nationality, gender and age do not significantly influence choice behavior.

While the resulting marginal effect for “*Paymenteffort*” is very intuitive, the result for the certainty equivalent is puzzling. For each cent the certainty equivalent of a subject is higher (i.e. the less risk-averse or the more risk-loving she is), the probability of the lottery being chosen *decreases* by 0.0004. This seems very small at first sight, and notably, the negative influence of an equally higher payment earned in the effort task is about four times as large (the marginal effect is -0.0019). However, the standard deviation of the certainty equivalent is about 214. Therefore, going up one standard deviation from the average subject corresponds to a decline of the probability of the lottery being chosen by 8.56 percent. This is a substantial and insofar a fairly unexpected result as it calls the current method of using certainty equivalents to explain actual risk behavior into question when describing choices between a risky lottery and a certain amount for that the decision maker has to exert effort to earn it.

Summing up our results so far, the behavior that we observe in the treatment group is best described as follows. After being informed about the opportunity to choose the risky payment option after the completion of the slider task, the subjects exert as much effort as they can<sup>67</sup> and make an ordinary ex post decision between the lottery and the (at that point) certain payoff earned in the task. To make money very easily by intentionally exerting either no effort or significantly less effort than the subjects in the control group and engaging in pure gambling instead is not a prevailing motive, even though we know that there is also a reverse causation in the sense that anticipated lottery choices affect effort exertion. This effect is

---

<sup>67</sup> Maybe they could, in principle, do more, but they choose to exert only a limited amount of effort. However, the point is that we observe the same effort levels as for their counterparts in the control group.

either very small, or there are only very few test persons who truly make an ex ante decision in favor of the lottery, and therefore it does not drive the results of the experiment in a way that a significant treatment effect on effort exertion would occur.

If the lottery choice is made ex post, it resembles an ordinary trade-off between a risky and a certain payment. Ex ante, however, “*Paymenteffort*” is not certain in a sense that the test persons cannot be perfectly sure about their effort level, at least if they do not decide to exert no effort at all. They need to learn about their skills in order to estimate correctly how well they can perform in the task, and therefore only the ones who are very pessimistic about their skills or who have a very high disutility from working on the task can be sufficiently confident in predicting that choosing the lottery will be the better alternative for them.

The following table reports the effort levels of the individuals in the treatment group over the stages and by the frequency of lottery choices. From the 37 subjects in this group, eight never choose the lottery, fourteen only choose it once, seven choose it twice and eight subjects always choose the risky payment option whenever it is available.

Table 4.5: Effort Levels by Stages and Frequency of Lottery Choices

	Frequency of lottery choices							
	0 (n=8)		1 (n=14)		2 (n=7)		3 (n=8)	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
<b>Effort S2</b>	9.75 <sub>a</sub>	2.21	9.21 <sub>a</sub>	1.15	9.00 <sub>a</sub>	2.34	6.88 <sub>a</sub>	1.56
<b>Effort S3</b>	15.63 <sub>a</sub>	1.10	14.36 <sub>a</sub>	0.99	14.29 <sub>a,b</sub>	1.67	9.37 <sub>b</sub>	2.25
<b>Effort S4</b>	20.50 <sub>a</sub>	1.73	17.29 <sub>a</sub>	0.83	17.14 <sub>a</sub>	1.24	6.88 <sub>b</sub>	2.29
<b>Effort S5</b>	22.63 <sub>a</sub>	1.29	19.86 <sub>a</sub>	0.73	20.71 <sub>a</sub>	1.54	7.63 <sub>b</sub>	2.76
<b>Effort S6</b>	23.00 <sub>a</sub>	1.39	21.21 <sub>a</sub>	0.85	22.86 <sub>a</sub>	2.45	21.38 <sub>a</sub>	1.65

Note: Values in the same row not sharing the same subscript are significantly different at the 90% confidence level in the two-sided test of equality for column means. Tests are adjusted for all pairwise comparisons within a row using the Bonferroni correction.

There are no statistically significant differences between effort levels of all subjects who either never choose the lottery or those who choose it only once or twice. However, we see that the subjects who always choose the lottery (i.e. in stages 3, 4, and 5), exert significantly less effort than the other test persons in these periods. Even though the difference is insignificant, it seems like these are the persons with the weakest performance in the first

round of the slider task, i.e. they are probably the ones who are (rightfully) the most pessimistic about their own skills. Note that the effort levels in the first three groups (zero, one, and two lottery choices) increase steadily over the stages, while the subjects who always choose the risky payment option exert even less effort in the fourth and fifth stage than they do in the third stage. We interpret this as a clear signal that these few subjects are the only ones who decide *ex ante* to exert very little effort, as they are going to choose the lottery option anyways. All other subjects constantly increase their performance over the course of the experiment and we can therefore say that if they make a choice in favor of the lottery, this decision is made *ex post*. Another interesting conclusion that can be drawn from this table is that obviously, intentional risk-taking does not have any detrimental long-term effects on the subjects' performance. In the last round of the slider task, when the lottery is no longer available, the effort level of the last subsample (with three choices in favor of the lottery) almost triples to catch up the difference to all other groups completely. One could have expected that the persons who continuously worked on the task with constantly ameliorating results are better trained in the last round and able to achieve higher scores than the ones who purposely exerted little effort beforehand, but this is clearly not the case.

#### 4.3.3 Risk Preferences and Lottery Choices

As mentioned before, the lottery is frequently chosen and the choice behavior is surprisingly different from what the risk preferences that were elicited by the certainty equivalents would predict. We would expect a subject who has earned an amount lower than her certainty equivalent to choose the risky payment option and vice versa. In the three rounds in which the lottery is available in the experimental group, we observe in total 111 of these choices. The following contingency table describes the relationship between the choice of the two payment options and the value constellation of the certainty equivalent and the payment that has been earned in the slider task.

Table 4.6: Value Constellations between Certainty Equivalents and Earnings from Slider Task versus Choices of Payment Mode

		Choselottery = 0	Choselottery = 1	Total
<b>CE ≤ payment effort task</b>	Count	22	1	23
	Expected Count	12.2	10.8	23.0
	CE – payment from task	-151	-66	
	Std. Residual	2.8***	-3.0***	
<b>CE &gt; payment effort task</b>	Count	37	51	88
	Expected Count	46.8	41.2	88.0
	CE – payment from task	270	289	
	Std. Residual	-1.4	1.5	
<b>Total</b>	Count	59	52	111
	Expected Count	59.0	52.0	111

Note: “Count” reports actual frequencies. “Expected Count” reports expected frequencies if value constellation and lottery choice were independent. “CE – payment from task” reports the average differences between certainty equivalents and payments earned in the slider task. “Std. Residual” reports the standardized residual between actual and expected frequencies. \*\*\*, \*\*, \* indicate statistical significance at the 99%, 95%, and 90% confidence levels.

In 23 of 111 cases, the subjects achieve a sufficient score in the slider task to earn more than their certainty equivalent. Consequentially, in almost all choices (22 out of 23) the payment earned in the task is preferred over the lottery. Unsurprisingly, the actual count of lottery choices is significantly different at the 99% confidence level from the expected count, i.e. if the choices were independent of the value constellation. On the other hand, in 88 cases the subjects only earn a lower amount. With the same degree of consistency in risk preferences, we would then expect that the lottery is virtually always the preferred payment option. However, only 51 out of 88 choices are in favor of the risky payment, while 37 choices are inconsistent with the elicited risk preferences. These numbers are not statistically significantly different from the expected count figures. In fact, when testing for an underlying binomial distribution with a probability of 0.5, we cannot reject that lottery choice is determined by a simple coin toss when the certainty equivalent is higher than the payment from the slider task ( $p\text{-value} = 0.165$ ). Overall, a Pearson  $\chi^2$ -Test indicates that value constellation and lottery choice are not independent, i.e. the subjects choose the payment option that is more beneficial to them ( $\chi^2 = 21.043$ ,  $p < 0.001$ ). However, we have shown that this result is driven by the minority of cases where the subjects have earned an amount in the slider task that exceeds their individual certainty equivalent.

When looking at the actual monetary values for the cases where the lottery is chosen and where it is not, it becomes clear that the average differences (in cents) between the individual certainty equivalents and the payments from the effort task are not statistically significantly different from each other for the two subsamples ( $t = 0.62$ ,  $p = 0.539$ ). When subjects choose the sure payoff, their valuation of the lottery is on average 2.70 Euros higher, compared to 2.89 Euros when they actually choose the lottery. Thus, we cannot argue that the cases where the payment from the effort task is chosen despite contrary risk preferences are those for which the differences in monetary values are less pronounced.

The choice of the payment mode can only be predicted accurately by the elicited certainty equivalents if the amount earned in the slider task is high enough to exceed the valuation of the lottery. Thus, we see that the trade-off between risky and certain payoffs yields different outcomes depending on the fact whether earning the certain payoff is costless (as in the choice lists used to elicit the certainty equivalents) or associated with effort (as in the slider task) in a way that a substantial fraction of subjects is willing to accept a much lower certain payment in the latter case. Note that effort costs should push decisions in the opposite direction: the relative attractiveness of the lottery *increases* if the utility of the certain payoff is diminished by costs.

In principle, there are two hypotheses that could explain this pattern. First, the certainty equivalents that are elicited in the first stage are too high to reflect the actual choice behavior in the later rounds of the experiment. Second, the certain payoffs from the slider task are valued higher than the ones from the choice list. While the first hypothesis is more of a methodological issue, the second one would make up a behavioral explanation for our observations. While each of these hypotheses could explain our results independently, it might also be the case that both are true to a certain extent. Within our experimental framework, we are not able to disentangle the two and more research is needed on this issue.

#### 4.3.4 Beliefs

To see whether the participants expect different effort levels in the two experimental groups, we perform a pooled regression similar to the one that is reported in Table 4.3 with the reported belief as the dependent variable instead of the effort level.<sup>68</sup> There is no significant treatment effect. Even though the beliefs are on average a little bit higher in the control

---

<sup>68</sup> The regression table is available from the authors upon request.

group<sup>69</sup>, the regression reveals that this is in fact a gender effect. The proportion of male participants is higher in the treatment group, and men have significantly lower beliefs about their peers' effort levels than women. Thus, the subjects do not expect a significant extent of shirking (i.e. low effort provision) when the alternative risky payment option is available. Given that their own behavior is the subjects' only source of information to base their beliefs on, it seems plausible that the beliefs are in line with the actual behavior.

#### **4.4 Discussion**

In our experiment we find that risk-taking is very common, but not in order to avoid effort. The only subjects who exert considerably less effort are the ones who always make use of the risky lottery – and they seem to be the ones with the lowest base productivity anyway. Thus, we might see a “resignation effect” here. In general, we conclude that risk-taking is the result of an ex post decision. Choice behavior in these decisions should be determined by the relative attractiveness – the assigned utility – of the two alternatives, namely receiving the amount that has been earned in the slider task or playing the lottery. Interestingly, we observe less risk-taking here than we would have expected on the basis of the risk preferences that were elicited through the certainty equivalents at the beginning of the experiment: when subjects fail to earn a payment that is higher than their certainty equivalent in the slider task, lottery choices could as well be determined by a coin toss. In principle, there could be two explanations for this. The first one is that the subjects might appreciate the monetary outcome of their own work more than a payment that “falls from sky”, like it is the case in the choice lists to elicit the certainty equivalents. This is different from enjoyment of the task, which we will discuss later. Choosing the payment from the task even if the certainty equivalent is higher could then be motivated by the wish not to let go the merits of one's hard work. The second explanation could be that the certainty equivalents are not measured properly. It could be the case that the participants show a tendency to switch over from the lottery to the certain payment somewhere in the center of the list, which would, due to the skewness of the lottery, result in risk-seeking preferences which then turn out not to be accurate in later choices between the lottery and the payment from the slider task. However, we have to point out that the choice lists are properly incentivized, so we should not simply assume that the subjects did not take them serious. Even though we cannot pin down the exact reason, we present

---

<sup>69</sup> Summary statistics are available from the authors upon request.



evidence that choice behavior is more risk averse (or less risk-seeking) when real effort is associated with earning the certain payoff. More research is needed to identify behavioral foundations and methodological limitations that may play a role here.

To recapitulate the main result once again, “making easy money” is not a prevailing motive for risk-taking. However, if there is the possibility to take a risk and this option is sufficiently attractive, it is of course still tempting. The appeal of the lottery should, in an *ex post* decision, be independent from the specific payment scheme that is underlying the accomplished task. Still, as a suggestion for future research, it would be interesting to see whether risk-taking and effort provision differ under, for example, fixed payments, linear variable payment schemes, or tournaments. As in our experiment, careful calibration is essential to produce viable results: if the lottery is more attractive to everybody, no matter what the effort levels and the risk preferences are, it will always be chosen. We conducted pre-tests to find a suitable balance, so that approximately half – 52 out of 111 – of the choices that were made in the experimental group are in favor of the risky outside option. Another variation of the experimental setup that could produce interesting results is to incorporate immediate feedback on the results of the lottery, i.e. to investigate the role of the so-called “house money effect” (see Johnson and Thaler (1990)) in a real effort experiment. In our experiment, the subjects are only paid in the very end for only one randomly determined stage and they do not know whether the outcome of the lottery was favorable or not when they enter the next stage. It could be that, if the test persons received immediate feedback, we might observe path-dependent effort exertion in a way that those who “successfully” played the lottery become more likely to choose the risk-taking option again in later rounds and that these later decisions are truly *ex ante* decisions that negatively affect effort.

We use the slider task developed by Gill et al. (2011), because it produces a wide range of outcomes, offers high learning opportunities, is easy to understand but not easy to do in limited time, and can be repeated without suffering from a shortfall of many other real effort tasks, namely that subjects remember results from former rounds and reproduce them without actually exerting effort. This is likely to happen when solving the same math problems or solving the same mazes over and over again. If, on the other hand, the tasks differ in each round, it is not a repeated measurement anymore. A potential problem of every real effort task that we cannot rule out with certainty is enjoyment of the task. If the test persons work because it is fun, there are no costs of effort exertion in the form of disutility, but even benefits from working. Bhattacharya and Dugar (2012) compare effort levels in two different

tasks under equal incentives and find higher effort in the task that is more interesting for the subjects. Dickinson (1999) points out that both “on the job” and “off the job” leisure should be taken into account in classical labor / leisure decision problems. We cannot explicitly measure the effort costs associated with working on the slider task, but we do not see any particular reason why the task should be perceived as enjoyable by the subjects. Furthermore, the slider task is, even though it is still fairly new, already established in the literature, most prominently in Gill and Prowse (2012).

Finally, we would like to point out that the artificial separation between effort provision and gambling in our experiment is unrealistic, but necessary to answer our research question: in reality, gambling may still look like the result of effort provision. For a manager, effort provision translates into thoroughness to make optimal decisions. Think about two asset managers who both have the same amount of funds to invest. Manager A, who exerts more effort to identify the optimal asset allocation, will still effectively invest the same amount of money like Manager B, who just throws darts at the “Financial Markets” section of her newspaper. Taking risk, in reality, is therefore much more a hidden action problem than in our experiment, where choices and effort provision are perfectly observable. Furthermore, managers also bear the additional risk of losing their job that should be very important when making the effort provision / risk-taking decision. Risk-taking can also increase or decrease the probability of keeping one’s job.

## Appendix to Chapter 4

### Appendix 4.A / General Instructions

Thank you for your participation in this experiment! Please remain seated over the whole course of the experiment and do not communicate with the other participants. You may ask questions to the instructor at any time by raising your hand.

In this experiment you can earn points by working on a number of tasks and by predicting the performance of the other participants in the experiment in these tasks. All participants will work on exactly the same tasks during the experiment. The way points can be earned in the different tasks is explained before each task.

- a) At the end of the experiment one task will be selected randomly and you will receive the amount of money that you earned in this task, depending on your choices and your performance (in the decision making task, one choice will be selected randomly, see outline below).
- b) Additionally, one task (not necessarily the same as above) will be selected randomly and you will receive a payment if you predicted the performance of the other participants in this task correctly (only applies to slider task rounds 2 – 5, see outline below).
- c) For completing the experiment you will receive an additional show up fee of 7 Euros regardless of your points score.

Decision Making Task	Slider Task Round 1	Slider Task Round 2	Slider Task Round 3	Slider Task Round 4	Slider Task Round 5
a) Make choices between fixed payments and a gamble	a) Your own performance in this round	a) Your own performance in this round b) Predict performance of the other participants in this round	a) Your own performance in this round b) Predict performance of the other participants in this round	a) Your own performance in this round b) Predict performance of the other participants in this round	a) Your own performance in this round b) Predict performance of the other participants in this round

## 190

For each choice in the list, please indicate whether you prefer "Left" or "Right". If you have made all choices, please press the "Confirm" button:

Appendix 4.C / Certainty Equivalent Elicitation, Choice List 2

Periode

1 von 6

In the list below you have to choose between two options. "Left" and "Right". Option "Left" is a gamble with two possible outcomes: with a probability of 70% you receive 0 cents and with a probability of 30% you receive 1250 cents. The gamble is the same in each choice (i.e. in each row of the list). Option "Right" is a fixed amount of money that is different in each row of the list. Please choose row by row whether you prefer to play the gamble or to receive the fixed amount instead. One row will be selected randomly and you will either receive the fixed payment for that row or play the lottery, according to the choice you made. Note that you have to switch from "Left" to "Right" exactly once.

Choice Number	Left	Right	Your Choice
1	Gamble	450 cents	Left <input type="radio"/> Right <input type="radio"/>
2	Gamble	452 cents	Left <input type="radio"/> Right <input type="radio"/>
3	Gamble	454 cents	Left <input type="radio"/> Right <input type="radio"/>
4	Gamble	456 cents	Left <input type="radio"/> Right <input type="radio"/>
5	Gamble	458 cents	Left <input type="radio"/> Right <input type="radio"/>
6	Gamble	460 cents	Left <input type="radio"/> Right <input type="radio"/>
7	Gamble	462 cents	Left <input type="radio"/> Right <input type="radio"/>
8	Gamble	464 cents	Left <input type="radio"/> Right <input type="radio"/>
9	Gamble	466 cents	Left <input type="radio"/> Right <input type="radio"/>
10	Gamble	468 cents	Left <input type="radio"/> Right <input type="radio"/>
11	Gamble	470 cents	Left <input type="radio"/> Right <input type="radio"/>
12	Gamble	472 cents	Left <input type="radio"/> Right <input type="radio"/>
13	Gamble	474 cents	Left <input type="radio"/> Right <input type="radio"/>
14	Gamble	476 cents	Left <input type="radio"/> Right <input type="radio"/>
15	Gamble	478 cents	Left <input type="radio"/> Right <input type="radio"/>
16	Gamble	480 cents	Left <input type="radio"/> Right <input type="radio"/>
17	Gamble	482 cents	Left <input type="radio"/> Right <input type="radio"/>
18	Gamble	484 cents	Left <input type="radio"/> Right <input type="radio"/>
19	Gamble	486 cents	Left <input type="radio"/> Right <input type="radio"/>
20	Gamble	488 cents	Left <input type="radio"/> Right <input type="radio"/>
21	Gamble	490 cents	Left <input type="radio"/> Right <input type="radio"/>
22	Gamble	492 cents	Left <input type="radio"/> Right <input type="radio"/>
23	Gamble	494 cents	Left <input type="radio"/> Right <input type="radio"/>
24	Gamble	496 cents	Left <input type="radio"/> Right <input type="radio"/>
25	Gamble	498 cents	Left <input type="radio"/> Right <input type="radio"/>
26	Gamble	500 cents	Left <input type="radio"/> Right <input type="radio"/>

For each choice in the list, please indicate whether you prefer "Left" or "Right". If you have made all choices, please press the "Confirm" button:

Confirm

## Appendix 4.D / Instructions Slider Task (1 of 5)

Periode	2 von 6
---------	---------

In this round, you will work on the "slider task". To earn points, you have to position as many sliders as you can to the middle position. Your result will determine your payment for this round. The number of sliders that you position exactly in the middle will be squared. The result equals to your payment in cents. A slider in the middle position shows a value of 50. There are in total 48 sliders on the screen. You can reposition each slider as often as you want. **Please have a look at the printed payment scheme to see the payment for each number of correctly positioned sliders.**

The "slider task" will be repeated five times throughout this experiment. In the first round you have 105 seconds to position as many sliders as you can. **In each subsequent round, you will have 8 seconds more than before to work on the task.**

To move the sliders, you are only allowed to use the mouse. IT IS FORBIDDEN TO USE THE KEYBOARD DURING THE SLIDER TASK! If you use the keyboard during the slider task, you will be expelled from the experiment and you will not earn any money.

If you do not want to continue working on the slider task, you can press the "Skip" button at any time and enjoy some cartoons. You maintain the points score that you have achieved so far.

After pressing the OK button below, you will see some screenshots of the slider task that will help you to familiarize with the structure of the screen. After this small tutorial, a countdown of three seconds will appear before the task starts.

OK

## Appendix 4.E / Instructions Slider Task (2 of 5)

[illegible]

## Appendix 4.F / Instructions Slider Task (3 of 5)

Periode

2 von 6

Tutorial Screen Slider Task 2 of 4

Next

1 von 5

Verbleibende Zeit [sec]: 66

3

Currently, your points score is:

Stop

**Your current score is displayed here. Three sliders show a value of 50, while one slider has a value of 52 and all other sliders are still positioned at 0. Thus, the current points score is 3**

50

50

50

52

0

0



### Appendix 4.G / Instructions Slider Task (4 of 5)

[illegible]

## 196

196

## Appendix 4.I / Payment Scheme Slider Task

Number of correctly positioned sliders	Your payment in cents
1	1
2	4
3	9
4	16
5	25
6	36
7	49
8	64
9	81
10	100
11	121
12	144
13	169
14	196
15	225
16	256
17	289
18	324
19	361
20	400
21	441
22	484
23	529
24	576
25	625
26	676
27	729
28	784
29	841
30	900
31	961
32	1024
33	1089
34	1156
35	1225
36	1296
37	1369
38	1444
39	1521
40	1600
41	1681
42	1764
43	1849
44	1936
45	2025
46	2116
47	2209
48	2304

## Appendix 4.J / Belief Elicitation

Periode

3 von 6

Before you start the slider task, please state the **average number of sliders** that you think will be positioned correctly **by all other participants** of the experiment (excluding yourself) in this round. Note that your own payment will only depend on your own performance in the slider task (if you do not choose the gamble instead).

If the actual average performance of all other participants of the experiment (excluding yourself) is within the range of your prediction  $+1$  /  $-1$ , you will receive a payment of 500 cents for this task. Otherwise, you will receive 0 cents.

Please enter the number in the field below and press the "Proceed to task" button afterwards. A countdown of three seconds will appear before the task starts.

The number of correctly positioned sliders in this round will be on average:

Proceed to task

## Appendix 4.K / Instructions Slider Task with Choice of Payment Mode

Periods	3 von 6
---------	---------

In the next round, the slider task will be repeated. **This time, you may choose between the payment that you earn in the slider task and a gamble.** You will make your choice after the completion of this round.

In this round, you have 113 seconds to position as many sliders as you can (8 seconds more than in the round before).  
To move the sliders, you are only allowed to use the mouse. IT IS FORBIDDEN TO USE THE KEYBOARD DURING THE SLIDER TASK! If you use the keyboard during the slider task, you will be expelled from the experiment and you will not earn any money.  
If you do not want to continue working on the slider task, you can press the "Skip" button at any time and enjoy some cartoons. You maintain the points score that you have achieved so far.

**Payment from slider task:** the number of sliders that you position exactly in the middle will be squared. The result equals to your payment in cents. Please have a look at the printed payment scheme to see the payment for each number of correctly positioned sliders.

**Gamble:** The gamble has two potential outcomes. With a probability of 70 percent, you will receive a payment of 0 cents. With a probability of 30 percent, you will receive a payment of 1250 cents. This is the same gamble as the one from the beginning of the experiment. You will not receive any feedback on the outcome of the gamble until the end of the experiment.

OK

Appendix 4.L / Ex Post Choice of Payment Mode

Periode

4 von 6

You can now choose whether you would like to receive the **payment of 529 cents** that you have just earned in the slider task or to **play a gamble instead**.  
If you play the gamble, you will receive **0 cents with a probability of 70 percent** and **1250 cents with a probability of 30 percent**. This is the same gamble as the one from the beginning of the experiment.  
You will not receive any feedback on the outcome of the gamble until the end of the experiment.

☐ Payment from slider task

☐ Gamble

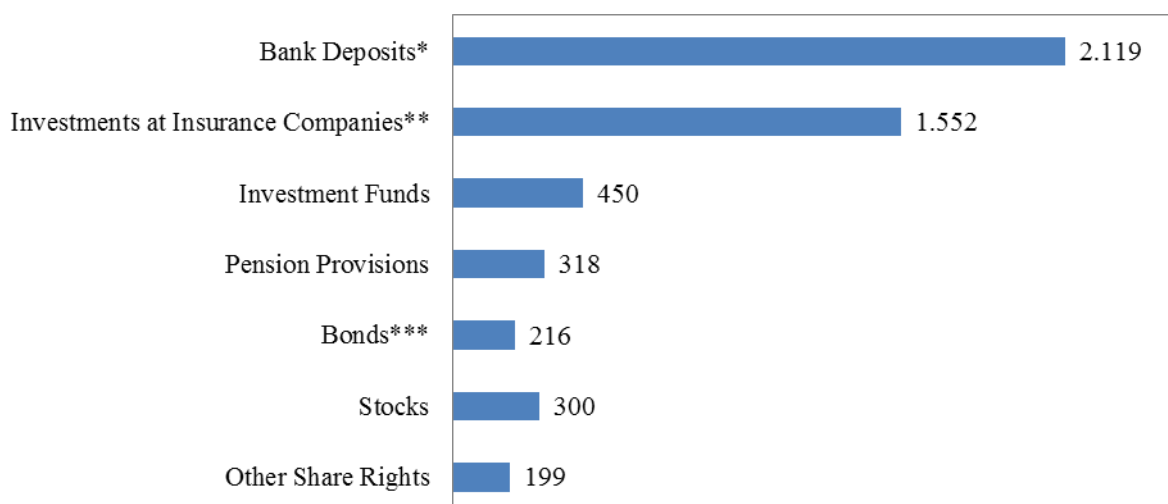
Confirm

## 5 Final Remarks

We discussed in the introduction of this dissertation that in experimental research there are wide opportunities to control the environment in which decisions are made and behavior can be observed. This is advantageous to measure latent constructs like preferences or types of behavior that are unobservable in the field, but it also means that experimental setups are often abstract and artificial so that the obtained results are not necessarily directly transferable to the real world.

For example, in the KID experiment we observe an overall high tendency to invest in investment funds while only a small portion of the money is put on the checking account (16.4 percent across all treatments). Figure 5.1 on the other hand shows the distribution of financial assets in German households in the year 2013, and we see that bank deposits in fact have the largest share with in total 41 percent of all financial assets, and notably assets held as bank deposits are almost five times as much as in investment funds. This demonstrates that at least the absolute figures that we obtain from our experiment should not be taken for granted. More important are the observed tendencies to invest more or less under a certain experimental condition.

Figure 5.1: Financial Assets of Private Households in Germany 2013 in Billion Euros



Note: \* includes cash, demand funds, time deposits, savings deposits, savings certificates, and other accounts receivable, \*\* includes pension funds, \*\*\* includes money market funds. Source: BVR (2014), p. 8.

The comparison of results from different experiments is also potentially problematic. An elementary phenomenon that has been discovered in experimental research is the existence of framing effects. Behavior is dependent on the environmental context and especially on the way decision problems are described. The most prominent example is the “Asian Disease Problem” where different ways of stating the expected outcomes of available countermeasures to fight the disease produce different decisions, even though the outcomes are numerically equivalent in terms of the number of cured and killed people. Another example relates directly to the different approaches that are commonly used to elicit risk preferences, namely to infer them from choices between lotteries or from monetary valuations. Lichtensein and Slovic (1971) are the first to present experimental evidence on such preference reversals. There are two binary lotteries with identical expected values, but the first one has a relatively high probability of winning a moderate payoff while the second one offers a small chance of winning a very high payoff. When the test persons are asked to choose which of the lotteries they would rather play, the first lottery is chosen by most people. When, on the other hand, the test persons shall place a bid for each of the lotteries, the second lottery usually produces higher valuations. The authors conclude that “choices tend to be determined by probabilities, while bids are most influenced by payoffs” (p. 47). Hence, choice lists and the BDM procedure – both are used in the experiments presented in this thesis – are likely to produce different certainty equivalents.

The applied measuring methods, the formulation of the provided information and questions, and the way behavioral incentives are installed, exert influence on the obtained results – and potentially in an unexpected or unforeseeable way. In principle, all these factors are controllable, but they can – and usually do – differ substantially among experiments and from real-world circumstances. Hence, the results that are presented in the single chapters of this dissertation should be recognized separately and independently. Especially in experimental research, there is a distinct culture of replication and gradual variation of earlier studies for exactly these reasons, where failure to replicate earlier results or significant changes due to punctual modifications of the experimental setups demonstrate context sensitivity and limitations in the generalization of the results.



# Bibliography

Aldy, Joseph E. and W. Kip Viscusi (2003), *The Value of a Statistical Life: A Critical Review of Market Estimates Throughout the World*, The Journal of Risk and Uncertainty, Vol. 27, No. 1, p. 5-76.

Allais, Maurice (1953), *Le comportement de l'homme rationnel devant le risque, critique des postulats et axiomes de l'école Américaine*, Econometrica, Vol. 21, No. 4, p. 503-46.

Anderson, Lisa R. and Jennifer M. Mellor (2008), *Predicting Health Behaviors with an Experimental Measure of Risk Preference*, Journal of Health Economics, Vol. 27, p. 1260–1274.

Andreoni, James and Charles Sprenger (2010), *Certain and Uncertain Utility: The Allais Paradox and Five Decision Theory Phenomena*, University of California at San Diego, Working Paper, available at <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.153.7812&rep=rep1&type=pdf>.

Arrow, Kenneth J. (1963), *Social Choice and Individual Values*, 2<sup>nd</sup> edition, Wiley, New York, NY.

Aumann, Robert J., Sergiu Hart and Motty Perry (2005), *Conditioning and the Sure-Thing Principle*, Center for the Study of Rationality, The Hebrew University of Jerusalem, Jerusalem, Discussion Paper 393, available at <http://www.ma.huji.ac.il/hart/papers/stp.pdf>.

Azrieli, Yaron, Christopher P. Chambers, and Paul J. Healy (2012), *Incentives in Experiments: A Theoretical Analysis*, Working Paper, available at [http://healy.econ.ohio-state.edu/papers/Azrieli\\_Cambers\\_Healy-CombinedDecisions.pdf](http://healy.econ.ohio-state.edu/papers/Azrieli_Cambers_Healy-CombinedDecisions.pdf).

Bär, Michaela, Alexander Kempf, and Stefan Ruenzi (2011), *Is a Team Different from the Sum of its Parts? Evidence from Mutual Fund Managers*, Review of Finance, Vol. 15, No. 2, p. 359-396.

Barber, Brad M., Chip Heath, and Terrance Odean (2003), *Good Reasons Sell: Reason-Based Choice among Group and Individual Investors in the Stock Market*, Management Science, Vol. 49, No. 12, p. 1636-1652.

Becker, Gordon M., Morris H. DeGroot and Jacob Marschak (1964), *Measuring Utility by a Single-Response Sequential Method*, Behavioral Science, Vol. 9, p. 226-232.

Bem, Daryl J., Nathan Kogan, and Michael A. Wallach (1964), *Diffusion of responsibility and level of risk taking in groups*, Journal of Abnormal and Social Psychology, Vol. 68, No. 3, p. 263-274.

Bhattacharya, Haimanti and Subhasish Dugar (2012), *Status Incentives and Performance*, Managerial and Decision Economics, Vol. 33, p. 549-563.

Binmore, Ken (1994), *Playing Fair: Game Theory and the Social Contract I*, MIT Press, Cambridge, MA.

Binmore, Ken (2009), *Rational Decisions*, Princeton University Press, Princeton, NJ.

Bone, John, John Hey, and John Suckling (1999), *Are Groups More (or Less) Consistent Than Individuals?*, Journal of Risk and Uncertainty, Vol. 8, p. 63-81.

Breakwell, Glynis M. (2007), *The Psychology of Risk*, Cambridge University Press, Cambridge, UK.

Brown, Keith C., W.V. Harlow and Laura T. Starks (1996), *Of Tournaments and Temptations: An Analysis of Managerial Incentives in the Mutual Fund Industry*, Journal of Finance, Vol. 51, No. 2, p. 85-110.

Budescu, David, Enrico Diecidue, and Steffen Keck (2012), *Group Decisions Under Ambiguity: Convergence to Neutrality*, INSEAD Working Paper No. 2012/43/DS, available at SSRN: <http://ssrn.com/abstract=2040655>.

Busse, J. A. (2001), *Another Look at Mutual Fund Tournaments*, Journal of Financial and Quantitative Analysis, Vol. 36, No. 1, p. 53-73.

BVR (2014), *BVR – Volkswirtschaft special*, Bundesverband der Deutschen Volksbanken und Raiffeisenbanken (ed.), Berlin, available at [http://www.bvr.de/p.nsf/B93A8874C0AF75EEC1257C440042DDEA/\\$FILE/BVRVolkswirtschaftSpecial.pdf](http://www.bvr.de/p.nsf/B93A8874C0AF75EEC1257C440042DDEA/$FILE/BVRVolkswirtschaftSpecial.pdf).

CESR (2010), *CESR's guidelines on the methodology for the calculation of the synthetic risk and reward indicator in the Key Investor Information Document*, Reference Number CESR/10-673, available at [http://www.esma.europa.eu/system/files/10\\_673.pdf](http://www.esma.europa.eu/system/files/10_673.pdf).

CESR (2010a), *CESR's template for the Key Investor Information Document*, Reference Number CESR/10-794, available at [http://www.esma.europa.eu/system/files/10\\_794.pdf](http://www.esma.europa.eu/system/files/10_794.pdf).

Charness, Gary, Edi Karni, and Dan Levin (2013), *Ambiguity attitudes and social interactions: An experimental investigation*, Journal of Risk and Uncertainty, Vol. 46, p. 1-25.

Condorcet, Jean-Antoine-Nicolas de Caritat, marquis de (1785), *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Imprimerie royale, Paris.

Davis, James H. and Verlin B. Hinsz (1982), *Current Research Problems in Group Performance and Group Dynamics*, in Herman Brandstatter, James H. Davis, and Gisela Stocker-Kraichgauer (eds.), Group Decision Making, London, Academic Press, p. 1-22.

- Dickinson, David L. (1999), *An Experimental Examination of Labor Supply and Work Intensities*, Journal of Labor Economics, Vol. 17, No. 4, pt.1, p. 638-670.
- Dohmen, Thomas and Armin Falk (2011), *Performance Pay and Multi-Dimensional Sorting: Productivity, Preferences, and Gender*, American Economic Review, Vol. 101, p. 556-590.
- Eliasz, Kfir, Debraj Ray, and Ronny Razin (2006), *Choice Shifts in Groups: A Decision-Theoretic Basis*, The American Economic Review, Vol. 96, No. 4, p. 1321-1332.
- Ellsberg, Daniel (1961), *Risk, Ambiguity, and the Savage Axioms*, The Quarterly Journal of Economics, Vol. 75, No. 4, p. 643-669.
- Eriksson, Tor and Marie Claire Villeval (2008), *Performance-Pay, Sorting, and Social Motivation*, Journal of Economic Behavior & Organization, Vol. 68, p. 412-421.
- Fischbacher, Urs (2007), *z-Tree: Zurich Toolbox for Ready-made Economic Experiments*, Experimental Economics, Vol. 10, No. 2, p. 171-178.
- French, K. and J. Poterba (1991), *Investor Diversification and International Equity Markets*, American Economic Review, Vol. 81, No. 2, p. 222-226.
- Gill, David and Victoria Prowse (2011), *A novel computerized real effort task based on sliders*, IZA Discussion Paper 5801, available at [https://faculty.cit.cornell.edu/vlp33/paper\\_effort\\_task\\_291009.pdf](https://faculty.cit.cornell.edu/vlp33/paper_effort_task_291009.pdf).
- Gill, David and Victoria Prowse (2012), *A Structural Analysis of Disappointment Aversion in a Real Effort Competition*, American Economic Review, Vol. 102, No. 1, p. 469-503.
- Gill, David, Victoria Prowse and Michael Vlassopoulos (2013), *Cheating in the Workplace: An Experimental Study of the Impact of Bonuses and Productivity*, Working Paper, available at SSRN: <http://ssrn.com/abstract=2109698>.
- Greenwald, Anthony G. (1976), *Within-Subject Designs: To Use or not to Use*. Psychological Bulletin, Vol. 83, No. 2, p. 314-320.
- Halevy, Yoram (2007), *Ellsberg Revisited: An Experimental Study*, Econometrica, Vol. 75, No. 2, p. 503-536.
- Haß, Lars H., Denis Schweizer and Maximilian Troßbach (2013), *Rational Decisions, Skills, and Group Dynamics under Risk and Ambiguity*, unpublished working paper.
- Heinberg, Aileen, Angela Y. Hung and Joanne K. Yoong (2010), *Do Risk Disclosures Affect Investment Choice?*, RAND Working Paper No. WR-788, available at SSRN: <http://ssrn.com/abstract=1688038>.

Huber, Joel, Wesley A. Magat, and W. Kip Viscusi (1987), *An Investigation of the Rationality of Consumer Valuations of Multiple Health Risks*, The RAND Journal of Economics, Vol. 18, No. 4, p. 465-479.

Hylland, Aanund and Richard Zeckhauser (1979), *The Impossibility of Bayesian Group Decision Making with Separate Aggregation of Beliefs and Values*, Econometrica, Vol. 47, No. 6, p. 1321-1336.

Johanning, Lutz and Maximilian Troßbach (2014), *Risk Disclosure in Key Investor Information Documents, Risk Perception, and Investment Decisions*, unpublished working paper.

Johnson, Eric J. and Richard H. Thaler (1990), *Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice*, Management Science, Vol. 36, No. 6, p. 643-660.

Kahneman, Daniel and Amos Tversky (1979), *Prospect Theory: An Analysis of Decision under Risk*, Econometrica, Vol. 47, No. 2, p. 263-292.

Kahneman, Daniel and Amos Tversky (1984), *Choices, values, and frames*, American Psychologist, Vol. 39, p. 341-50.

Keller, L. Robin, Rakesh K. Sarin, and Jayavel Sounderpandian (2007), *An Examination of Ambiguity Aversion: Are Two Heads Better than One?*, Judgment and Decision Making, Vol. 2, No. 5, p. 390-397.

Lejuez, Carl W., Jennifer P. Read, Christopher W. Kahler, Jerry B. Richards, Susan E. Ramsey, Gregory L. Stuart, David R. Strong, and Richard A. Brown (2002), *Evaluation of a Behavioral Measure of Risk Taking: The Balloon Analogue Risk Task (BART)*, Journal of Experimental Psychology: Applied, Vol. 8, p. 75-84.

Lichtenstein, Sarah and Paul Slovic (1971), *Reversal of Preferences between Bids and Choices in Gambling Decisions*, Journal of Experimental Psychology, Vol. 89, p. 46-55.

Markowitz, Harry M. (1952), *Portfolio Selection*, Journal of Finance, Vol. 7, p. 77-91.

Mas-Colell, Andreu, Michael D. Whinston and Jerry R. Green (1995), *Microeconomic Theory*, 1<sup>st</sup> edition, Oxford University Press, New York, NY.

Maskin, E. and J. Riley (1984), *Optimal Auctions with Risk-Averse Buyers*, Econometrica, Vol. 52, No. 6, p. 1473-1518.

Maskin, E. and J. Riley (2000), *Equilibrium in Sealed High Bid Auctions*, Review of Economic Studies, Vol. 67, p. 439-454.

Mathauschek, Barbara, Bettina Rockenbach, and Abdolkarim Sadrieh (2007), *Teams take the better risks*, Journal of Economic Behavior & Organization, Vol. 63, p. 412-422.

OLG Stuttgart (2013), Court Decision December 18, 2013, reference number 9 U 52/13, OpenJur 2014, 8402, available at <http://openjur.de/u/679966.html>.

Riley, J. and W.F. Samuelson (1981), *Optimal Auctions*, American Economic Review, Vol. 71, No. 3, p. 381-392.

Ross, Stephen A. (2004), *Compensation, Incentives, and the Duality of Risk Aversion and Riskiness*, The Journal of Finance, Vol. 59, No. 1, p. 207-225.

Samuelson, Paul (1947), *Foundations of Economic Analysis*, Harvard University Press, Cambridge, MA.

Samuelson, Paul (1952), *Probability, Utility, and the Independence Axiom*, Econometrica, Vol. 20, No. 4, p. 670-678.

Savage, Leonard J. (1954), *The Foundations of Statistics*, John Wiley & Sons, New York, NY.

Schlenker, Barry A. and Michael F. Weigold (1991), *Accountability and Risk Taking*, Personality and Social Psychology Bulletin, Vol. 17, No. 1, p. 25-29.

Slovic, Paul and Amos Tversky (1974), *Who Accepts Savage's Axiom?*, Behavioral Science, Vol. 19, No. 6, p. 368-373.

Stoner, James A.F. (1968), *Risky and cautious shifts in group decisions: The influence of widely held values*, Journal of Experimental Social Psychology, Vol. 4, No. 4, p. 442-459.

Trautmann, Stefan, Maximilian Troßbach, and Gijs van de Kuilen (2013), *Making Easy Money – Do We Trade Risk for Laziness?*, unpublished working paper.

Trautmann, Stefan and Gijs van de Kuilen (2014), *Ambiguity Attitudes*, in: Keren, Gideon and George Wu (eds.), *The Blackwell Handbook of Judgment and Decision Making*, Blackwell Publishing, Oxford (forthcoming).

Trautmann, Stefan, Ferdinand Vieider, and Peter Wakker (2008), *Causes of Ambiguity Aversion: Known versus Unknown Preferences*, Journal of Risk and Uncertainty, Vol. 36, p. 225-243.

Trautmann, Stefan and Ferdinand Vieider (2011), *Social Influences on Risk Attitudes: Applications in Economics*, in: S. Roeser (ed.), *Handbook of Risk Theory*, Springer.

Trautmann, Stefan and Richard Zeckhauser (2013), *Shunning Uncertainty: The neglect of Learning Opportunities*, Games and Economic Behavior, Vol. 79, p. 44-55.

Tversky, Amos (1996), *Contrasting Rational and Psychological Principles of Choice*, in Richard L. Zeckhauser, Ralph L. Keeney, and James K. Sebenius (eds.), *Wise Choices: Decisions, Games, and Negotiations*, Harvard Business School Press.

van Dijk, Frans, Joep Sonnemans and Frans van Winden (2001), *Incentive Systems in a Real Effort Experiment*, European Economic Review, Vol. 45, p. 187-214.

Vickrey, W.S. (1961), *Counterspeculation, Auctions and Competitive Sealed Tenders*, Journal of Finance, Vol. 16, p. 8-37.

Vieider, Ferdinand M. (2011), *Separating real incentives and accountability*, Experimental Economics, Vol. 14, p. 507-518.

von Neumann, John and Oskar Morgenstern (1944), *The Theory of Games and Economic Behavior*, 1<sup>st</sup> edition, Princeton University Press, Princeton, NJ.