Four Contributions to Experimental Economics

Inaugural-Dissertation
zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften
durch die
Rechts- und Staatswissenschaftliche Fakultät
der Rheinischen Friedrich-Wilhelms-Universität
Bonn

vorgelegt von
Daniel Wiesen
aus Rheinbach

Bonn 2011

Dekan: Prof. Dr. Klaus Sandmann
Erstreferent: Prof. Dr. Dr. h.c. mult. Reinhard Selten
Zweitreferent: Prof. Dr. Armin Falk

Tag der mündlichen Prüfung: 06.05.2011

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn
(http://hss.ulb.uni-bonn.de/diss_online) elektronisch publiziert.

Contents

List of Figures

List of Tables

Introduction

Over the last thirty years findings from economic experiments substantially contributed to a better understanding of a wide variety of phenomena in different branches of economics (Plott and Smith 2008). In areas like industrial organization, game theory, public choice or labor economics controlled laboratory experiments became commonplace.

In contrast, in health economics the use of laboratory experimentation is rather in its infant stages. This is somewhat surprising as prominent proponents, like the US health economist Victor R. Fuchs, have already argued that incorporating methods of experimental economics into health economic research might lead to great benefits for the latter (Fuchs 2000).

Similarly, very little experimental research has focussed on individual risk attitudes of higher orders, like prudence and temperance[1], so far. Various experimental methods have been developed to investigate risk aversion (e.g., Holt and Laury 2002) and to test theories of decision-making under risk (e.g., Camerer 1989, Hey and Orme 1994). It is, thus, surprising that an appropriate method to test for higher-order risks is still lacking.

The four experimental studies presented in the dissertation at hand aim to fill this gap in the two respective research areas. The first two chapters present novel experimental methods to explore individual attitudes towards higher-order risks. In the third chapter, a laboratory experiment is introduced in order to study the influence of payment incentives on physician behavior. The final chapter analyzes the link between other-regarding motivations and physician payment incentives for two different subject pools. The following paragraphs provide some background and briefly introduce the four chapters of the dissertation.

Risk and uncertainty are important in numerous economic decision situations. Consequently, understanding individual risk attitudes is closely related to the purpose of pre-

[1]Within the expected utility framework, assuming differentiability of a utility function u, prudence and temperance are defined as $u''' > 0$ and $u^{(4)} < 0$, respectively.

dicting economic behavior (Dohmen et al. forthcoming b). Indeed, it is well known that risk aversion only partially captures individual risk attitudes (e.g., Gollier 2001). Numerous important behavioral traits, like precautionary savings and preventive behavior, are rather driven by higher-order risk preferences such as prudence or temperance. In particular, prudence exhibits a broad range of important implications on economic behavior within the expected utility framework. In his seminal article, Kimball (1990) shows that prudence is necessary and sufficient for a precautionary savings motive. Moreover, prudence has been shown, for example, to play an important role for precautionary bidding in auctions (Esö and White 2004) and insurance demand (Fei and Schlesinger 2008). However, the predictions derived from prudence might be nothing more than interesting mind games if one cannot test for prudence empirically.

So far, only very few papers have looked at the empirical support for prudence, mostly via the precautionary demand for savings (e.g., Dynan 1993, Carroll and Kimball 2008). In a first attempt to use methods of experimental economics, Tarazona-Gomez (2004) found weak evidence for prudence relying on an expected utility setting.

In a recent article, Eeckhoudt and Schlesinger (2006) showed, outside expected utility theory, equivalence of prudence to a preference over a simple lottery pair that is appropriate for experimental testing. In their definition, a prudent individual prefers—given two equally likely future states—to have an unavoidable zero-mean risk in the state where her wealth is higher. Put differently, prudence shows a type of preference for the disaggregation of two harms—a sure loss and a zero-mean risk.

A novel experimental method *to test* individuals for prudence, employing the lotteries of Eeckhoudt and Schlesinger, is presented in Chapter 1.[2] Moreover, it compares prudence with a preference for a high third moment, referred to as skewness-seeking. The methodology allows for a rather general implementation of Eeckhoudt and Schlesinger's prudence lotteries. In the experiment, subjects are asked to make choices on prudence lottery pairs. The new ballot box representation of the compound lotteries employed in this study could find application in other experiments on decision-making under risk. To contrast prudence with skewness seeking, subjects are also asked to make choices on lottery pairs testing for a skewness preference.

The only other experimental study testing for prudence, employing Eeckhoudt and Schlesinger's lotteries, has been introduced by Deck and Schlesinger (2010) who find evi-

[2]Chapter 1 is based on the joint working paper with Sebastian Ebert entitled "*An Experimental Methodology Testing for Prudence and Third-order Preferences*", Bonn Econ Discussion Paper No. 2009/21.

dence for prudence. Behavioral data of the experiment in Chapter 1 also evidence prudence on the aggregate as well as at the individual level. 65% of subjects' choices are prudent. 47% of individuals are classified as prudent whereas only 8% are imprudent. Although the experimental presentation and the lotteries employed are different from Deck and Schlesinger, the proportion of prudent choices in the experiment is similar to their finding of 61%. The first significant extension to their result is that behavioral data indicate more prudent decisions when the zero-mean risks are left-skewed, indicating also that kurtosis is influential on subjects' choices. Another major finding is that prudence is not sufficiently well captured by skewness-seeking.

The purpose of the experiment introduced in Chapter 2 is *to measure* risk aversion, prudence and temperance, and to compare their relative importance.[3] Notice that similar to prudence, Eeckhoudt and Schlesinger (2006) define temperance as a preference for the disaggregation of two independent zero-mean risks. Whereas the experiment in Chapter 1 and that of Deck and Schlesinger test for risk attitudes in a yes-or-no fashion, this chapter investigates the intensity of risk preferences. To the best of our knowledge no empirical comparison of the strength of the different risk attitudes exists and, thus, the experimental study presented in Chapter 2 marks the first step in this direction. The experimental approach is a combination of the method introduced in Chapter 1 and the multiple price list format which is popular to measure risk aversion (e.g., Holt and Laury 2002). The experimental method, outside the expected utility framework, elicits premia that make individuals indifferent between the risk averse (prudent, temperate) and risk loving (imprudent, intemperate) choice. The lotteries used in the experiment were calibrated such that premia are comparable. The main result is somewhat surprising. Namely, the desired compensation for prudence is significantly higher than that for risk aversion, which in turn is significantly higher than that for temperance. This may suggest that prudence has received not enough attention compared to risk aversion, which itself and its implications are investigated both theoretically and empirically in numerous papers. Moreover, we observe that the well documented gender effect that women are more risk averse (Croson and Gneezy 2009) extends to risks of higher order.

A crucial issue in health economics research is to understand how incentives from payment systems affect providers of medical services, i.e., physicians (McGuire 2000). In particular,

[3]The experimental study presented in Chapter 2 is a joint work with Sebastian Ebert entitled "*Joint Measurement of Risk Aversion, Prudence and Temperance*".

to achieve socially desirable outcomes payment incentives must be structured appropriately. However, in most health systems fee-based payments are the dominant form to pay physicians. Recently, lump-sum payments per patient have been introduced as an alternative form of physicians' remuneration. From a theoretical point of view, fee-for-service payment schemes may encourage physicians to provide too many medical services whereas the opposite incentive, namely to provide too few services, is embedded in lump-sum capitation payments (Ellis and McGuire 1986, Newhouse 2002). Empirical evidence on the effect of payment incentives on physician provision behavior is rather mixed, although an influence in general can be acknowledged (e.g., Dumont et al. 2008). Moreover, field studies face various difficulties, like, for example, multiple and unobservable influences on physicians' decisions or country-specific payment system variations, that make the generalization of results rather difficult (Gosden et al. 2001).

A controlled laboratory experiment, analyzing the influence of fee-for-service and capitation payments on physicians' provision behavior, is introduced in Chapter 3 to overcome these possible problems. Prospective physicians, i.e., medical students, decide on the quantity of medical service for several different patients.[4] Behavioral data indicate that patients are overserved under fee-for-service and underserved under capitation payments. However, financial incentives are not the only motivation for physicians' choices as the patient benefit shows to be of considerable importance as well. The patients are affected differently by the two payment systems. Those patients in need of a low level of medical services are better off under capitation, whereas patients with a high need for medical services gain a higher health benefit when physicians are paid by fee-for-service.

The experiment presented in Chapter 4 makes use of the experimental setup introduced in Chapter 3.[5] This chapter explores behavioral differences between prospective physicians and a 'standard' experimental subject pool, i.e., non-medical students. Experimental data show substantial differences across subject pools. Medical service choices of prospective physicians indicate a markedly lower tendency to overprovide (underprovide) patients under fee-for-service (capitation) than choices of non-medical students. This behavioral result suggests that financial incentives work only in an alleviated way for prospective physicians, as their behavior is mainly driven by patient-regarding (altruistic) motivations.

[4]Chapter 3 is based on a joint work with Heike Hennig-Schmidt and Reinhard Selten entitled "*How Payment Systems Affect Physicians' Provision Behavior – An Experimental Investigation*" available as Bonn Econ Discussion Paper No. 29/2009.

[5]Chapter 4: Hennig-Schmidt, H. and D. Wiesen (2010): "*Are Prospective Physicians Different? Evidence from an Experimental Study on a Payment System Variation.*"

Each chapter is based on a self-contained discussion paper. So readers can explore, in accordance to their interests, how to test for higher-order risk attitudes (Chapter 1) and how to measure them (Chapter 2), how incentives embedded in payment schemes affect physicians' behavior (Chapter 3) and how prospective physicians and 'standard' experimental subjects respond to incentives differently (Chapter 4).

Chapter 1

Testing for prudence and skewness seeking

1.1 Introduction

It is well known that risk aversion only partially captures an individual's risk preferences. An example is the following lottery pair defined by Mao (1970). Lottery M_A pays zero with a probability of $p = \frac{1}{4}$ and 2000 with the counter-probability of $\frac{3}{4}$. Lottery M_B pays 1000 with a probability of $\frac{3}{4}$ and 3000 with a probability of $\frac{1}{4}$. Statistically, these lotteries have the same mean and variance, but M_B is more skewed to the right. While M_A may seem 'riskier', the preference of M_B over M_A is not implied by risk aversion but by prudence.[1]

This was shown by Menezes et al. (1980) who illustrate that M_B exhibits more downside risk and who further characterize prudence as downside risk aversion. They also show that prudence, unlike risk aversion, relates to measures of skewness, in particular to the third central moment and semi-target variance.[2] Thus, prudence plays an important role when considering preference towards downside and left-skewed risks.

Such risks occur frequently in everyday life. For example, most insurance contracts address downside risks similar to M_A (where the arbitrary choice of p is much smaller than $\frac{1}{4}$ and refers to the insurance event). Similarly, on the gain side, M_B corresponds to the risk of a typical lottery ticket. The payoff structures of numerous assets exhibit downside risk. For example, the payoff distribution of a (defaultable) bond resembles M_A. Downside risk might be even more crucial both to investors and private individuals than

[1] Within the expected utility framework, risk aversion can be defined as $u'' < 0$ and prudence as $u''' > 0$.

[2] Chiu (2005) links prudence to a strong measure of skewness due to van Zwet (1964). See also Chiu (forthcoming) and Ebert (2010) for more on prudence and skewness.

comparable symmetric risks which are addressed by risk aversion. This is also evidenced by the intensive use of downside risk measures such as value-at-risk. In his seminal study, Mao surveyed business executives' reasons on investments of the type M_A and M_B.

Within the expected utility theory (EUT), numerous important implications of prudence on economic behavior have been shown. To name a few, Kimball (1990) coined the term prudence and showed that it is necessary and sufficient for a precautionary savings motive.[3] Eeckhoudt and Gollier (2005) analyze the impact of prudence on prevention, i.e., the action undertaken to reduce the probability of an adverse effect to occur.

These general concepts find application in various areas of economics and finance. The broad range of applications, provided in the following non-exhaustive list, emphasizes the importance of prudence. In health economics, Courbagé and Rey (2006) show that prudence is an important factor in preventive care decisions within a medical decision-making context. Esö and White (2004) show that there can be precautionary bidding in auctions when the value of the object is uncertain and when bidders are prudent. Likewise, White (2008) analyzes prudence in bargaining. Treich (forthcoming) shows that prudence can decrease rent-seeking efforts in a symmetric contest model. Fagart and Sinclair-Desgagné (2007) investigate prudence in a principal-agent model with applications to monitoring and optimal auditing. Within a macroeconomic consumption and labor model, Eeckhoudt and Schlesinger (2008) analyze the impact of prudence on policy decisions such as changes in the interest rate. Other examples are insurance demand (e.g., Fei and Schlesinger 2008) or life-cycle investment behavior (e.g., Gomes and Michaelides 2005). Even in environmental economics prudence plays a decisive role; Gollier (2010) finds an ecological prudence effect when discounting future environmental impacts.

Prudence is also necessary (but not sufficient) for decreasing absolute risk aversion, properness (Pratt and Zeckhauser 1987) and standard risk aversion (Kimball 1993). Further, prudence is exhibited by all the commonly used utility functions (Brockett and Golden 1987), in particular power and exponential utility. Thus, implicitly, prudence is assumed widely in the economics and finance literature.

While preference of M_B over M_A is necessary but not sufficient for prudence, Eeckhoudt and Schlesinger (2006) presented a more general lottery preference which is equivalent to prudence. Given two equally likely future states, a prudent individual prefers to have

[3]That means the awareness of uncertainty in future payoffs will raise an individual's optimal saving today. The relationship between precautionary savings and the third derivative of the utility function was already recognized by Leland (1968) and Sandmo (1970).

an unavoidable zero-mean risk in the state where her wealth is higher. Equivalently, she prefers to have the unavoidable harms of a sure loss and a zero-mean risk in different future states rather than in the same state. More generally, Eeckhoudt and Schlesinger define proper risk apportionment of all orders (where prudence corresponds to order 3). This new understanding of risk preferences does not rely on EUT. Further, it can be generalized to the multi-attribute case as shown in Eeckhoudt et al. (2007) or Tsetlin and Winkler (2009).

Despite the substantial amount of literature on prudence and the frequency of downside risk in general, there is little empirical, i.e., experimental, research on prudence. This is in sharp contrast to other theories of decision-making under risk. Theoretical predictions derived from prudence might be no more than an interesting mind game if we cannot test their validity. Currently the share between theoretical and empirical, in particular experimental ones, is extremely unbalanced.

Some empirical papers trace prudence via the precautionary savings motive relying on Kimball's (1990) EUT-model (e.g., Dynan 1993, Carrol 1994 and Carroll and Kimball 2008). To test the theories and behavioral traits based on prudence in a more controlled environment, we need a methodology to test individuals for prudence in the laboratory. The first attempt in this direction was made by Tarazona-Gomez (2004), who finds weak evidence for the existence of prudence. Her experiment relies on a certainty equivalent approach involving tabulated trinomial lotteries. It is based on strong assumptions and approximations within expected utility theory. The only other much more elegant approach to test for prudence is Deck and Schlesinger (2010). Using six pairs of Eeckhoudt and Schlesinger's lotteries, they find some evidence for prudence.

The contribution of this study is as follows. Firstly, we propose a method to test for prudence in a laboratory setting. Some results overlap with those of Deck and Schlesinger, whose research was started independently of ours. However, our methodology allows for a more general implementation of the zero-mean risks in the prudence lotteries. This feature is necessary, because on the theoretical side we show that prudence is not only a preference for high skewness—just as risk aversion is not only a preference for a low variance—but that this preference is robust towards different levels of kurtosis. This is referred to as the "kurtosis robustness feature of prudence." It is the zero-mean risk that drives the statistical properties (in particular the kurtosis) of the prudence lotteries. This makes them different from the simpler lotteries of Mao and from the ones of Deck and Schlesinger, who considered symmetric risks only. As a left-skewed zero-mean risk constitutes more harm to a prudent individual, one could conjecture a greater tendency to 'apportion the harms properly'. Indeed, in the experiment we observe significantly more prudent decisions when the risks to be apportioned are left-skewed. Although the experimental presentation and parameters are different from Deck and Schlesinger, our overall result that 65% of choices are prudent is close to their finding of 61%.

Secondly, we implement the lotteries, similar to Mao's survey, for the first time in an incentivized experiment. This is done to contrast the rather general Eeckhoudt and Schlesinger lotteries (ES lotteries), equivalent to prudence from simpler lotteries implying *skewness seeking* only. This applies to the lotteries used in the experiments of Tarazona-Gomez and Deck and Schlesinger. Skewness seeking can be motivated by the assumption of third order moment preferences.[4] Although moment preferences, in general, are incompatible with EUT, they are widely assumed in economic and financial modeling due to their simplicity and tractability.[5] Under *moment preferences*, individuals' decisions between two prospects only depend on the first few statistical moments of these prospects. When studying prudence only prospects with equal mean and variance will be compared. Such *third-order moment preferences* are equivalent to a preference for or against a high third central moment and refer to 'the' skewness of the prospect. That is, in this setting prudence is equivalent to skewness seeking. In the experiment the skewness seeking preference of M_B over M_A is more widely observed than preference over the prudence lotteries. There is also a significant positive correlation between the two and, consistent with theory, most individuals we diagnose as prudent prefer M_B over M_A. However, while skewness seeking has some approximative value in explaining preferences, for a more accurate test of prudence the ES lotteries should be implemented in a more general way. In particular, subjects do respond to differences in their kurtosis which leads us to reject the assumption of third-order moment preferences.

Thirdly, concerning the experimental methodology, we propose a novel graphical representation of compound lotteries in experiments. We also present a convenient method for lottery calibration in terms of their first three statistical moments. In particular, we analyze the statistical moments of the ES and Mao lotteries, where we also complement a

[4]Tarazona-Gomez (2004) explicitly makes this assumption. In particular, preferences are given by a utility function which is truncated at third order.

[5]For example, they underlie a large number of classical and also modern portfolio choice models, such as Kraus and Litzenberger (1976) or Briec et al. (2007). This is also why this theory is an active area of research. See, for example, Eichner (2008). Brockett and Kahane (1992) and Brockett and Garven (1998) show explicitly that they are incompatible with EUT.

result in Roger (forthcoming).

The chapter proceeds as follows. Section 2 analyzes the lotteries underlying the experiment, motivates the parameter choices and outlines the novel lottery calibration technique applied in this chapter. In Section 3, the research questions are stated. Section 4 describes the experimental design and procedure. In Section 5, results from the experiment are provided and Section 6 concludes.

1.2 Prudence and skewness seeking

In this section, we first define the lotteries employed in the experiment. Then we analyze and interpret their statistical properties and show how they relate to skewness seeking and prudence. The last subsection is concerned with the calibration of lottery parameters in the experiment.

1.2.1 Mao's and Eeckhoudt and Schlesinger's lotteries

We begin with the definition of binary lotteries in general.

Definition 1. *Let* $x_1, x_0 \in \mathbb{R}$, $x_1 > x_0$ *and* X *is a Bernoulli-distributed random variable with parameter* $p \in (0,1)$.[6] *A* (simple) binary lottery *denoted by* $L = L(p, x_1, x_0)$ *is defined as the random variable*

$$L = X \cdot x_1 + (1 - X) \cdot x_0.$$

In recognition of Mao (1970), we define the following class of lotteries and give an example in Figure 1.1.[7]

Definition 2. *Two binary lotteries* $L_X = L_X(p_X, x_1, x_0)$ *and* $L_Y = L_Y(p_Y, y_1, y_0)$ *constitute a Mao lottery pair or a Mao pair if they have equal mean and variances and* $p_X = 1 - p_Y$.

Intuitively, for a Mao lottery pair, if L_X has its high payoff associated with the *high* probability, then L_Y has its high outcome associated with the *small* probability, and vice versa.[8] Next we define the prudence lotteries of Eeckhoudt and Schlesinger (2006) and

[6] A Bernoulli-distributed random variable can take the two values 1 and 0 with probabilities $P(X = 1) = p$ and $P(X = 0) = 1 - p$, where $p \in (0, 1)$. Notice that it is convenient and mathematically tractable to use Bernoulli-distributed random variables for characterization of simple lotteries; see, e.g., Roger (forthcoming).

[7] Definition 2 specifies a class of lotteries that characterizes the risks analyzed in Mao's survey. Mao (1970) considered a concrete example of this class that motivated the paper of Menezes et al. (1980).

[8] This actually is how skewness manifests in a binary lottery; see our theorem in Subsection 2.3 and Ebert (2010) for a generalization to higher orders.

Figure 1.1: Example of a Mao lottery pair (M_A, M_B)

The lotteries above correspond to the Mao pair displayed to subjects in question MAO1 of the experiment. A Mao-preferent individual prefers lottery M_B (with a skewness of $+1.15$) over lottery M_A (with a skewness of -1.15).

give an example in Figure 1.2.

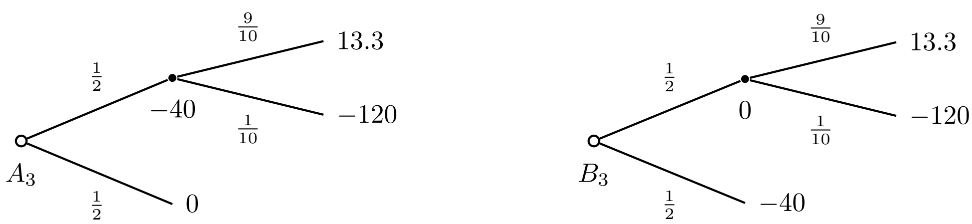
Definition 3 (Eeckhoudt-Schlesinger Prudence)**.** *Let X be a Bernoulli-distributed random variable with parameter $p = \frac{1}{2}$ and let $k > 0$. Let ϵ be a non-degenerate random variable independent of X with $\mathbb{E}[\epsilon] = 0$. Consider the lotteries*

$$A_3 = X \cdot (0) + (1 - X) \cdot (-k + \epsilon) \text{ and } B_3 = X \cdot (-k) + (1 - X) \cdot \epsilon.$$

These two lotteries as a pair are called (Eeckhoudt-Schlesinger) prudence lottery pair or ES pair. An individual is called prudent or said to have the ES preference if she prefers B_3 over A_3 for all values of k, for all random variables ϵ and for all wealth levels x (i.e., non-random variables $x \in \mathbb{R}_{>0}$).

Figure 1.2: Example of a prudence lottery pair (A_3, B_3)

Wealth level $x = 160$:



The lotteries above correspond to the ES pair displayed to subjects in question ES1 of the experiment. In the example, ϵ is left-skewed, implying that lottery A_3 has a larger kurtosis than lottery B_3 (see Proposition 3).

For the prudent option B_3 the additional zero-mean risk ϵ (i.e., the second lottery) occurs in the good state of the 50/50 gamble (i.e., in the state without the sure reduction in wealth, $-k$), whereas for the imprudent option A_3 the zero-mean risk occurs in the bad state. Intuitively, a prudent choice implies a 'disaggregation of harms' or 'proper risk apportionment'. Eeckhoudt and Schlesinger show that this preference is equivalent to prudence within EUT, i.e., $u''' > 0$. Menezes et al. (1980) define an increase in downside risk and show that prudence is equivalent to downside risk aversion. They reinterpret the results of Mao's survey and show that the lottery M_B has less downside risk than the corresponding lottery M_A.

Proposition 1 (Menezes, Geiss, Tressler, 1980)**.** *Let (M_A, M_B) denote a pair of Mao lotteries. Prudence is sufficient (but not necessary) for preferring M_B over M_A.*

It is interesting to note that Eeckhoudt and Schlesinger (2006) define proper risk apportionment of all orders n via an iterative nesting process of lotteries. For these lotteries B_n and A_n they then show that preferring B_n over A_n is equivalent to $sgn(u^{(n)}) = (-1)^{n+1}$ within EUT. Risk apportionment of order 4 is called temperance and there is also quite a large amount of theory concerned with the behavioral implications of temperance. Although in this study we focus on prudence, our methodology could be easily adapted to test for risk aversion and temperance. The experiment of Deck and Schlesinger (2010) actually finds some evidence of intemperate decisions.

1.2.2 Prudence, moments and skewness seeking

We now investigate the statistical features of the Mao and ES (prudence) lotteries in more detail. This will motivate the particular choices of the lottery pairs we implement in the experiment. If not noted otherwise, 'moments' refer to standardized central moments and the n^{th} moment of a random variable Z is given by $\mu_n^S(Z) := \mathbb{E}[(Z - \mathbb{E}[Z])^n] / (\mathbb{V}(Z))^{n/2}$. With $\nu(Z) := \mu_3^S(Z)$ and $\kappa(Z) := \mu_4^S(Z)$ we denote the third and fourth moment, respectively. All proofs are given in Appendix A.1.1.

Proposition 2 (Statistical Characterization of Mao lottery pairs)**.** *Consider a pair of Mao lotteries given by $L_X = L_X(p, x_1, x_0)$ and $L_Y = L_Y(1-p, y_1, y_0)$. Then for all $n \in \mathbb{N}$ we have*

$$\mu_n^S(L_X) = (-1)^n \mu_n^S(L_Y),$$

in particular

$$\nu(L_X) = -\nu(L_Y) \text{ and } \kappa(L_X) = \kappa(L_Y).$$

The following proposition generalizes Proposition 4 in Roger (forthcoming).

Proposition 3. *Consider an arbitrary Eeckhoudt-Schlesinger lottery pair in Definition 3.*

A_3 and B_3 have equal expectation and variance and thus $\mathbb{V}(A_3) = \mathbb{V}(B_3) =: \sigma^2$ is well-defined. Furthermore,

$$\begin{aligned}\nu(B_3) - \nu(A_3) &= \frac{3k\mathbb{E}[\epsilon^2]}{2\sigma^3} > 0 \text{ and} \\ \kappa(B_3) - \kappa(A_3) &= \frac{2k\mathbb{E}[\epsilon^3]}{\sigma^4} \text{ can be positive, negative, or zero.}\end{aligned}$$

The third and fourth moments, respectively, are sometimes referred to as 'the' skewness and 'the' kurtosis. However, there are numerous measures for these properties; see MacGillivray (1986) for an overview. In a recent paper, Ebert (2010) generalized our Proposition 3 and showed that $\mu_n^S(B_3) - \mu_n^S(A_3) > 0$ for all n odd. Further, for binary zero-mean risks such as those employed in our experiment, $\kappa(B_3) - \kappa(A_3) < 0 \iff \mu_n^S(B_3) - \mu_n^S(A_3) < 0$ for all n even.[9] Therefore, while thinking in third- and fourth-order terms will provide the reader with the correct intuition, our arguments actually apply to the very strong notions of skewness and kurtosis that refer to *all* odd and even moments, respectively.[10]

We know that Mao lotteries have equal mean and variance, and have the same kurtosis (see Proposition 2). We say that an individual is *skewness seeking* if she prefers the Mao lottery with the high (positive) skewness over the one with small (negative) skewness, which is well-defined as a consequence of Proposition 2.

Comparing Propositions 2 and Proposition 3, we see that both prudence and the Mao preference imply higher skewness to be beneficial to the individual. Unlike the Mao preference, prudence further requires that the lottery with the higher skewness is preferred no matter whether it has the smaller or higher kurtosis. That is, prudence implies a preference for skewness, but it also requires this preference to be robust towards variations in kurtosis. Ebert (2010) puts this observation on a more rigorous basis and refers to the "kurtosis robustness feature of prudence."

What is the origin of this extra of statistical structure of the ES lotteries compared to the Mao lotteries? From Proposition 3 we see that the prudent choice has the smaller kurtosis if and only if the zero-mean risk that has to be apportioned is left-skewed. The zero-mean risks of the lotteries employed in the experiment of Deck and Schlesinger (2010)

[9]The assumption of a binary risk is not crucial to his result. For example, the result also holds if the majority of odd moments of ϵ is negative; see Ebert (2010) for details.

[10]For random variables with a compact support, the sequence of (natural) moments characterizes the probability distribution. In the statistics literature, this is known as Hausdorff moment problem. In this sense, these descriptions of skewness and kurtosis for prudence lotteries are exhaustive.

were symmetric. This constantly implies the same kurtosis for the two prudence lotteries. Moreover, Roger (forthcoming) shows that the signs of all moments of ES lotteries with symmetric ϵ's coincide with those we derived in Proposition 2 for the Mao lotteries. Thus, from a statistical point of view, prudence lotteries with symmetric zero-mean risks are much closer to the skewness seeking lotteries of Mao than to the general proper risk apportionment lotteries of Eeckhoudt and Schlesinger (2006). Preference between the former lotteries is solely determined by skewness preference and does not reflect the kurtosis robustness feature of prudence.

In this study, we not only avoid this restriction, but also evaluate it. This requires a comprehensive experimental presentation of the compound ES lotteries as the skewed risks to be apportioned cannot be presented as a fair coin toss to subjects. In the experiment, subjects will also decide over Mao lotteries to test them for skewness seeking, which (theoretically) is necessary, but not sufficient to imply prudence.

1.2.3 Lottery calibration

In order to have the prudence and Mao lotteries in the same parameter range, Mao lottery pairs must be calibrated such that they are close to the prudence pairs.[11] We start with a theorem stating that a binary lottery with non-trivial variance and otherwise arbitrary first three moments always exists and the moments uniquely determine the lottery. It implies that every non-degenerate probability distribution with finite first three moments can be approximated up to the third moment by a binary lottery and this approximating lottery is unique. Binary lotteries are one of the main tools to examine decisions under risk and for testing associated theories like expected utility (see, e.g., Hey and Orme 1994) or prospect theory (Kahneman and Tversky 1979). Therefore, the following theorem might find application in many experiments and, in particular, is useful for calibration issues. The given equations conveniently allow to construct exactly the lottery an experimenter is looking for. Finally, the theorem gives intuition on how skewness manifests in binary lotteries; for more, see also Chiu (forthcoming) and Ebert (2010).

Theorem 1. *For constants $E \in \mathbb{R}$, $V \in \mathbb{R}_+^*$ and $S \in \mathbb{R}$ there exists exactly one binary lottery $L_X = L_X(p, x_1, x_0)$ such that $\mathbb{E}[L_X] = E$, $\mathbb{V}[L_X] = V$ and $\nu[L_X] = S$. Its parameters*

[11] This is a general issue in lottery choice experiments and has been shown to be important, for example, in the context of multiple price list formats to elicit risk preferences (see, e.g., Harrison and Rutström 2008). We will show that this calibration has an effect on subjects' decisions in Subsection 1.5.4.

are given by

$$p = \begin{cases} \frac{4+S^2+\sqrt{S^4+4S^2}}{8+2S^2} & \text{if } S < 0 \\ \frac{1}{2} & \text{if } S = 0 \\ \frac{4+S^2-\sqrt{S^4+4S^2}}{8+2S^2} & \text{if } S > 0 \end{cases},$$

$$x_1 = E + \sqrt{\frac{V \cdot (1-p)}{p}}, \quad x_0 = E - \sqrt{\frac{V \cdot p}{1-p}}.$$

Now we use this theorem to calibrate the Mao and ES pairs to each other. The Mao pair in Figure 1.1 and the ES pair in Figure 1.2 are an example. All four lotteries depicted have equal mean and variance and the differences in skewness between the ES pair and the Mao pair are also equal. When an ES pair and a Mao pair are calibrated to each other in this way, we call them *corresponding* lottery pairs. The following proposition gives an existence and uniqueness result for such a calibration.

Proposition 4 (Calibration). *Consider a prudence lottery pair* (A, B) *with finite first three moments. For every* $S > 0$ *there exists exactly one Mao lottery pair* (M_A, M_B) *such that*

$$\mathbb{E}[M_A] = \mathbb{E}[A] \text{ and } \mathbb{E}[M_B] = \mathbb{E}[B],$$
$$\mathbb{V}[M_A] = \mathbb{V}[A] \text{ and } \mathbb{V}[M_B] = \mathbb{V}[B] \text{ as well as}$$
$$\nu[M_A] = -S \text{ and } \nu[M_B] = S.$$

For $S = 0.5(\nu[B] - \nu[A])$ *the difference in third moments of the prudence pair equals the difference in third moments of the Mao pair and the quadratic error* $\Delta := (\nu[B] - \nu[M_B])^2 + (\nu[A] - \nu[M_A])^2$ *is minimized.*

1.3 Research questions

We propose a method to test for prudence, employ it in an experiment and test it for robustness. We put particular focus on whether prudence boils down to skewness seeking or if, on the other hand, we find evidence for the kurtosis robustness feature of prudence. This is explained in more detail in this section.

As explained in Section 2, the Mao lotteries are similar in structure to the ones employed in earlier studies (e.g., Tarazona-Gomez 2004) and differ by their skewness, but not by their kurtosis. Thus they help us to investigate the relationship of prudence and skewness seeking.

Research Question 1. *What is the relationship between prudence and the Mao preference?*

Answering research question 1 is of particular interest for several reasons. If prudence and the Mao preference are equivalent, skewness seeking seems to characterize prudence sufficiently well. Moreover, this would support moment preferences up to order three in general. If Mao-preferent individuals do not exhibit prudence, this implies that prudence is a stronger property, not only in theory, but also in practice. In particular, it is not sufficient then to ask lotteries based on the first three moments to test for prudence. Further, as shown in Subsection 1.2.3, then no binary lottery can be sufficiently complex to test for prudence.

Eeckhoudt and Schlesinger's definition of prudence (Definition 3) is very broad in scope. That is, the lottery preference must hold for any random variable ϵ, any loss $-k$, any wealth level x and, of course, is robust towards framing of the decision task. This is the reason why a 'simple binary lottery preference' can be equivalent to signing the derivatives of the utility function—looking more closely, the lottery preference is not that simple. In particular, the fact that the zero-mean risks are arbitrary adds a large amount of stochastic freedom to these lotteries. We will test in a systematic way which of these features do significantly influence subjects' decisions.

Concerning the robustness towards framing, we test whether it makes a difference if the task is to add the zero-mean risk ϵ or the fixed amount $-k$ to a state of the 50/50 gamble, given that the other item ($-k$ or ϵ, respectively) is already present in one state. This relates to the intuition of Eeckhoudt and Schlesinger's definition of prudence as 'proper risk apportionment'. Further, Definition 3 of prudence could be adapted such that the loss $-k$ is replaced by fixed gain in wealth k. The prudent choice then is the one where k and ϵ appear in the *same* state (a prudent individual prefers the unavoidable additional risk when wealth is higher). Further in-depth explanations are provided in Section 1.4. In short, we state the following research questions.[12]

Research Question 2. *Are individuals' decisions independent of whether the fixed amount k corresponds to a gain or a loss?*

[12]Research Questions 2 to 4 have been addressed to some degree in Deck and Schlesinger (2010). We will compare results in Section 1.5

Research Question 3. *Are individuals' decisions influenced by the wealth level x?*

Research Question 4. *Are individuals' decisions influenced by different framing of the decision task—whether they are asked to add the zero-mean risk ϵ or the fixed amount k to a state of the 50/50 gamble?*

Research Question 5. *Are individuals' decisions influenced by the skewness of the zero-mean risk ϵ and, therewith, the kurtosis of the prudence lotteries?*

We also will investigate whether the calibration of lotteries in terms of the first three moments (described in Subsection 1.2.3) does have an effect on subjects' decisions. Further, we will investigate whether age, gender and risk aversion have an impact on prudence.

1.4 Experimental design and procedure

The computerized experiment, programmed in z-Tree (Fischbacher 2007), comprises the three stages ES, MAO and RIAV. In total, each subject makes 34 individual choices over lottery pairs. The lottery outcomes are disclosed in Taler, our experimental currency. One Taler is worth €0.15 (about $0.20). Decisions are incentivized by a random-choice payment technique. That means, one out of 34 decisions is randomly drawn to determine solely a subject's payoff.[13] The lottery chosen by the individual in the randomly determined decision is actually played out at the end of the experiment.

In stage ES, we test subjects for prudence. Subjects decide over Mao pairs in stage MAO. In stage RIAV, we determine subjects' degree of risk aversion, employing the well-established method by Holt and Laury (2002). A questionnaire comprising demographic questions follows the experiment. We now describe the experimental stages in more detail.

[13]In economic experiments, it has become quite common to elicit a series of choices from participants and then to pay for only one selected at random; see Baltussen et al. (2010) for a fine overview. The random choice payment technique enables the researcher to observe a large number of individual decisions for a given research budget. However, the important question arises whether subjects behave as if each of these choices involves the stated payoffs. This issue has been analyzed, among various other setups, in experiments with pairwise lottery choice problems similar to our experiment. For example, Starmer and Sugden (1991) found clear evidence that under random payment subjects isolate choices as if paid for each task. Similar evidence was reported by Beattie and Loomes (1997) and Cubitt et al. (1998). In a lottery experiment with a multiple price list format and low incentives, Laury (2005) reports no significant difference in choices between paying for 1 or all 10 decision.

1.4.1 Prudence test embedded in a factorial design – Stage ES

In stage ES, we test whether individuals are prudent according to Definition 3. To this end subjects are asked to make preference choices over the 16 ES pairs $ES1, ES2, \ldots,$ ES16. We introduce a new ballot box representation to display the compound lotteries of the ES pairs. Figure 1.3 shows, as an example, how question ES1 (that has already been illustrated more formally in Figure 1.2) appears on subjects' decision screens. It must be understood as follows: Option A and Option B are displayed in the left and right panel of Figure 1.3, respectively. For both options the 50/50 gamble is depicted as a ballot box that contains two balls labeled "Up" and "Down". The displays of both Option A and Option B themselves are spatially separated, each into an upper panel containing the "Up-ball", and into a lower panel containing the "Down-ball". Now consider Option A. If the draw from the first ballot box is "Up", then the subject incurs a loss of 40 Taler and a second lottery (the zero-mean risk ϵ) follows. The zero-mean risk ϵ is also displayed in a ballot box format with 10 balls in total. Balls implying a loss (here: -120 Taler) are colored in yellow on subjects' decision screens and balls implying a gain (here: 13.3 Taler) are colored in white. In situation "Down" no second lottery follows and no loss occurs. For Option B, if the draw from the first ballot box is "Up", no loss occurs and a second lottery follows (the same ϵ as depicted in Option A). If the draw is "Down", a loss of 40 Taler occurs. The order of subjects' decision screens occurred for each subject in a different randomized order and also the position of the prudent option being either left or right on the screen has been randomized.

The ballot box representation interlinks decisions on the computer screen with the lottery play at the end of the experiment; see Figure 1.4. Further, it visualizes asymmetric zero-mean risks and all probabilities in an intuitive way.

To test Research Questions 2 to 5, we employ a completely randomized factorial design.[14] The factors are as follows: sign of k (Factor A), wealth level x (Factor B), framing (Factor C) and composition of ϵ (Factor D); see columns 6 to 9 in Table 1.1 for a complete design layout.

Along the illustration in Figure 1.3 we now explain how the factors of the factorial design translate into subjects' decision screens. When Factor A is at its low level ($k_1 = 40$), the outcomes of the 50/50 gamble are 0 Taler and -40 Taler. That is, the fixed amount added corresponds to a loss. Hence, in the example, the imprudent choice is Option A, as

[14] For a detailed description of the factorial design technique, see, e.g., Montgomery (2005).

Figure 1.3: Example of the lottery display in stage ES (Question ES1)

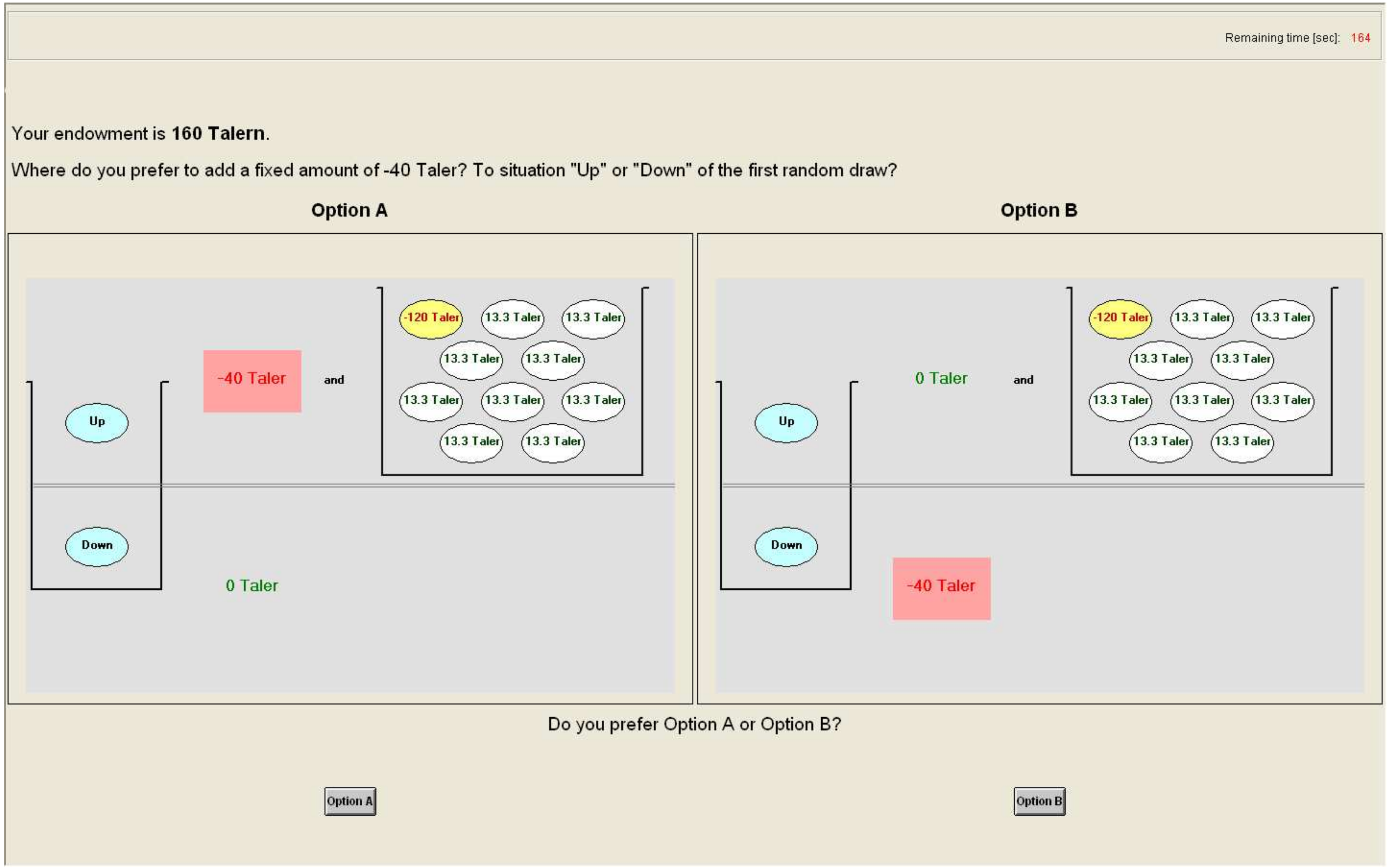


Figure 1.4: Sample of ballot boxes

This photograph shows an example of the ballot boxes used to determine subjects' payoffs at the end of the experiment from a decision made in stage ES, e.g., ES1 (compare to screenshot in Figure 1.3).

the additional zero-mean risk occurs in the bad state. At the alternative level of Factor A ($k_2 = -40$) the amount 40 is added, which corresponds to a gain and is displayed as a green bill on subjects' screens. With Factor A we test for an experimental framing effect (Research Question 2) and whether individuals really exhibit the intuition of proper risk apportionment. For example, if a subject consistently prefers the option where ϵ is added to outcome 0 Taler (independent of the sign of k) we could conjecture that this is due to framing and conclude that 0 is a so-called focal point.

Factor B tests for a wealth effect according to Research Question 3 and comprises the levels $x_1 = 160$ or $x_2 = 80$ Taler. This test is limited in that wealth levels are presented as endowments to subjects that they receive in order to accommodate possible negative lottery outcomes. However, in some tasks, the wealth effect can actually be quite substantial.[15] The wealth level on subjects' screens is indicated in the upper left corner. In Figure 1.3, it is set to 160 Taler.

In the example, the decision between the imprudent Option A and the prudent Op-

[15]For example, when comparing possible outcomes of prudent and imprudent choices for questions ES4 ($x_1 = 160$) and ES6 ($x_2 = 80$), the difference between subjects' possible payoff varies between 80 and 100 Taler, i.e., €12 to €15 (about $ 16 to $ 20).

tion B is whether in the up-state or in the down-state a fixed loss of 40 Taler is preferred *given* that the additional risk will be in the up-state. That is, the question on the decision screen is "*Where do you prefer to add a fixed amount of -40 Taler? To situation "Up" or "Down" of the first risky event?*" At the other level of Factor C, subjects are asked to which situation—either 0 or $-k$—of the 50/50 gamble to add another risky event (ϵ). Thus, the two levels of Factor C are "add k" (a sure reduction or increase in wealth) or "add ϵ" (a zero-mean random variable). Factor C directly relates to the intuition behind Eeckhoudt and Schlesinger's prudence definition of proper risk apportionment. It purely checks for a framing issue as the lotteries across levels of Factor C are identical in distribution.

With Factor D, we test if prudence is invariant under variation of the ϵ's (Research Question 5) or *equivalently*, for the kurtosis robustness feature of prudence. According to Proposition 3 and Ebert (2010), the prudent lottery choice B_3 has always the higher skewness compared to the imprudent choice A_3, i.e., $\mu_n(\epsilon) < 0$ for all n odd. It has the smaller kurtosis, i.e., $\mu_n(B_3) - \mu_n(A_3) < 0$ for all n even, if and only if ϵ is left-skewed. Thus, when varying the zero-mean risks, it is natural to vary their skewness systematically as this is the significant driver of the statistical differences between the prudence lotteries. The skewness of a binary lottery depends on its up-probability.[16] In our example, ϵ is left-skewed, such that the prudent lottery choice has the smaller kurtosis. If ϵ in the example had the signs of the outcomes switched it would be right-skewed and the prudent option had the higher kurtosis. As ϵ has a mean of zero, skewness has the following interpretation. A left-skewed ϵ yields a small gain with high probability and a large loss with a small probability. Further, as we display ϵ as a ballot box containing 10 balls, skewness translates one-to-one to the number of draws implying losses or gains, respectively. Indeed, in the example, ϵ implies a loss of 120 Taler with a 10% chance and a gain of 13.3 Taler with a 90% chance.

We denote the levels of Factor D "$\kappa(B_3) - \kappa(A_3) > 0$" (positive kurtosis difference) and "$\kappa(B_3) - \kappa(A_3) < 0$" (negative kurtosis difference). However, any of the mentioned equivalent interpretations (kurtosis difference, skewness of the zero-mean risk, composition of the ballot box) is captured by Factor D. These practical interpretations of kurtosis difference support our theoretical argument that restricting to symmetric ϵ's is a somewhat severe limitation for a procedure that aims to test for prudence.

To sum up, by specifying the four factors above, we manipulate the requirements in

[16]For this and the following arguments, see Theorem 1 and its proof in Appendix A.1.1.

Table 1.1: ES pairs with their underlying factors and their statistical properties

	ϵ				Factors				Statistical properties				
ES pair	p	z_1	$1-p$	z_0	A	B	C	D	$\mathbb{E}[A_3] = \mathbb{E}[B_3]$	$\mathbb{V}(A_3) = \mathbb{V}(B_3)$	$\nu(B_3) - \nu(A_3)$	$\kappa(B_3) - \kappa(A_3)$	Rel. freq.
ES1	0.90	13.33	0.10	−120.00	40	160	add $-k$	$\kappa(B_3) - \kappa(A_3) < 0$	−20.00	1,200.00	2.30	−9.48	0.6806
ES2	0.10	120.00	0.90	−13.33	40	160	add $-k$	$\kappa(B_3) - \kappa(A_3) > 0$	−20.00	1,200.00	2.30	9.48	0.5972
ES3	0.80	12.00	0.20	−48.00	40	160	add ϵ	$\kappa(B_3) - \kappa(A_3) < 0$	−20.00	688.00	1.92	−3.50	0.6111
ES4	0.20	48.00	0.80	−12.00	40	160	add ϵ	$\kappa(B_3) - \kappa(A_3) > 0$	−20.00	688.00	1.92	3.50	0.7083
ES5	0.70	12.00	0.30	−28.00	40	80	add $-k$	$\kappa(B_3) - \kappa(A_3) < 0$	−20.00	568.00	1.48	−1.33	0.5972
ES6	0.30	28.00	0.70	−12.00	40	80	add $-k$	$\kappa(B_3) - \kappa(A_3) > 0$	−20.00	568.00	1.48	1.33	0.6944
ES7	0.60	8.00	0.40	−12.00	40	80	add ϵ	$\kappa(B_3) - \kappa(A_3) < 0$	−20.00	448.00	0.60	−0.15	0.7222
ES8	0.40	12.00	0.60	−8.00	40	80	add ϵ	$\kappa(B_3) - \kappa(A_3) > 0$	−20.00	448.00	0.60	0.15	0.6944
ES9	0.90	13.33	0.10	−120.00	−40	160	add $-k$	$\kappa(B_3) - \kappa(A_3) > 0$	20.00	1,200.00	2.30	9.48	0.5000
ES10	0.10	120.00	0.90	−13.33	−40	160	add $-k$	$\kappa(B_3) - \kappa(A_3) < 0$	20.00	1,200.00	2.30	−9.48	0.7500
ES11	0.80	12.00	0.20	−48.00	−40	160	add ϵ	$\kappa(B_3) - \kappa(A_3) > 0$	20.00	688.00	1.92	3.50	0.6250
ES12	0.20	48.00	0.80	−12.00	−40	160	add ϵ	$\kappa(B_3) - \kappa(A_3) < 0$	20.00	688.00	1.92	−3.50	0.7083
ES13	0.70	12.00	0.30	−28.00	−40	80	add $-k$	$\kappa(B_3) - \kappa(A_3) > 0$	20.00	568.00	1.48	1.33	0.5139
ES14	0.30	28.00	0.70	−12.00	−40	80	add $-k$	$\kappa(B_3) - \kappa(A_3) < 0$	20.00	568.00	1.48	−1.33	0.6944
ES15	0.60	8.00	0.40	−12.00	−40	80	add ϵ	$\kappa(B_3) - \kappa(A_3) > 0$	20.00	448.00	0.60	0.15	0.5972
ES16	0.40	12.00	0.60	−8.00	−40	80	add ϵ	$\kappa(B_3) - \kappa(A_3) < 0$	20.00	448.00	0.60	−0.15	0.7222

This table describes the prudence lottery pairs $\mathrm{ES1}, \mathrm{ES2}, \ldots, \mathrm{ES16}$ in stage ES. ϵ is the zero-mean risk with its up-state z_1, its down-state z_0 and the respective probabilities p and $1-p$ shown in columns 2 to 5. The explicit arrangement of Factors A, B, C and D is given in columns 6 to 9. The columns 10 to 13 provide information of the difference in moments of the ES pairs and the last column shows the fraction of prudent choices observed for each lottery pair in the experiment that will be discussed in Section 1.5.3.

Eeckhoudt and Schlesinger's definition of prudence and for framing issues. We can test which factors have a severe impact on individuals' decisions such that they should be accounted for when testing for prudence. A complete overview of the 16 ES pairs, their statistical properties and the arrangement of factors is provided in Table 1.1.

1.4.2 Stage MAO

In this stage, we investigate whether subjects are Mao-preferent in order to answer Research Question 1. Applying Proposition 4, we obtain 8 different pairs of Mao lotteries, between which subjects have to state their preference. There are only 8 pairs, as the change in the kurtosis (Factor D) does not affect these lotteries (see Proposition 2). Thus lottery pair MAO1 corresponds to both lottery pairs ES1 and ES2, lottery pair MAO2 corresponds to ES3 and ES4, and so on. As the Mao lotteries imply negative outcomes, subjects are endowed with a certain amount of money equal to the wealth level x in the corresponding ES pairs. The Mao pairs are shown in Table 1.2.[17]

For the Mao lottery pairs we choose a graphical representation similar to the one proposed by Camerer (1989). An example of a decision screen can be found in the instructions to stage I in Appendix A.1.2.

1.4.3 Stage RIAV

In stage RIAV, we apply the well-known method of Holt and Laury (2002) to test for risk aversion; see the original article for details. We include their test in order to investigate the relationship between risk aversion and prudence.

1.4.4 Procedural details

The experiment was conducted at the *B*onn*E*con*L*ab. Overall 72 students of the University of Bonn from various fields participated in 9 experimental sessions in December 2008, January and February 2009. The stage order was varied systematically across sessions. Each session lasted for about 90 minutes. Subjects earned on average €18.50 (about $24.70).

The procedure of the experiment was as follows: firstly, experimenters extensively introduced the decision task and the entire procedure of the experiment to subjects. Secondly, before each experimental stage started, subjects were asked to answer control questions testing their understanding of the decision task. In particular, they were familiarized with the illustration of lotteries and their outcomes as well as probabilities. Only when subjects

[17] Analogous to stage ES, the order of subjects' decision screens is randomly permutated in stage MAO and the position of the Mao-preferent option is randomized.

Table 1.2: Mao pairs and their statistical properties

	M_A				M_B				Statistical properties		
Mao pair	p	x_1	$1-p$	x_0	p	y_1	$1-p$	y_0	$\mathbb{E}[M_A] = \mathbb{E}[M_B]$	$\mathbb{V}(M_A) = \mathbb{V}(M_B)$	$\nu(M_B) = -\nu(M_A)$
MAO1	0.75	0.00	0.25	−80.00	0.75	−40.00	0.25	40.00	−20.00	1200.00	1.15
MAO2	0.72	−3.48	0.28	−61.64	0.72	−36.52	0.28	21.64	−20.00	688.00	0.96
MAO3	0.67	−3.44	0.33	−54.30	0.67	−36.56	0.33	14.30	−20.00	568.00	0.74
MAO4	0.58	−1.81	0.42	−44.62	0.58	−38.19	0.42	4.62	−20.00	448.00	0.30
MAO5	0.75	40.00	0.25	−40.00	0.75	0.00	0.25	80.00	20.00	1200.00	1.15
MAO6	0.72	36.52	0.28	−21.64	0.72	3.48	0.28	61.64	20.00	688.00	0.96
MAO7	0.67	36.56	0.33	−14.30	0.67	3.44	0.33	54.30	20.00	568.00	0.74
MAO8	0.58	38.19	0.42	−4.62	0.58	1.81	0.42	44.62	20.00	448.00	0.30

This table shows the eight Mao pairs, i.e., MAO1, MAO2, ..., MAO8, employed in stage MAO. The final three columns provide information of the difference in moments of the Mao pairs.

had answered these questions correctly were they allowed to proceed to the decision stages of the experiment. Then, thirdly, subjects made the decisions in the experimental stages. Afterwards, subjects answered a questionnaire for which they received €4.00 ($5.34) in addition to their earnings from the experiment (comparable to a show-up fee). Finally, each subject's payoff was determined by a random-choice payment technique. To this end, for each subject one ball was chosen out of a set of balls numbered between 1 and 34 from a ballot box referring to a lottery pair from stage ES, MAO or RIAV. The subject's lottery choice in this randomly drawn lottery pair was then actually played out. In stages MAO and ES, the outcome was allocated to the subjects' wealth level in that decision, i.e., subjects could charge the coupon they obtained in the beginning. The ES lotteries were played out using ballot boxes resembling the lotteries displayed on subjects' decision screens (see the photograph in Figure 1.4). The binary lotteries in stages MAO and RIAV were played out using a ballot box with 100 balls numbered from 1 to 100. If, e.g., the up-state had a likelihood of 90%, a draw of the balls numbered $1, 2, \ldots, 90$ implied the corresponding up-payoff.

1.5 Experimental results

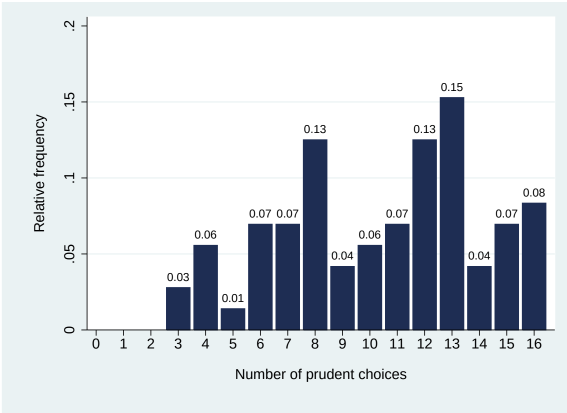
In this section we present the results of the experiment. We first describe subjects' choices at an aggregate level. Moreover, we report results from analyzing factors, the test of moment preferences and analysis of behavior at an individual level. Finally, we investigate the relationship between prudence and risk aversion as well as demographic characteristics.

1.5.1 Preliminary analysis

There is evidence for both prudence and Mao-preferent behavior at an aggregate level.[18] Figure 1.5 plots the relative frequencies of subjects' prudent choices. Overall, 65.10% of subjects' responses are prudent. This fraction is slightly higher than the 61% of prudent choices reported by Deck and Schlesinger (2010). In our sample, on average 10.42 of the choices are prudent with a standard deviation of 3.65. The median (mode) of prudent choices is 11 (13). The observed behavior in stage ES differs significantly from arbitrary behavior. Formally, we can reject the null hypothesis that the median of subjects' prudent

[18] To rule out possible stage order effects, we compare responses from sessions with stage order MAO-ES with responses from sessions with stage order ES-MAO. The null hypothesis that both samples are drawn from the same distribution cannot be rejected (for ES-responses: $p = 0.413$ and for MAO-responses: $p = 1.000$, two-sided two-sample Kolmogorov-Smirnov test).

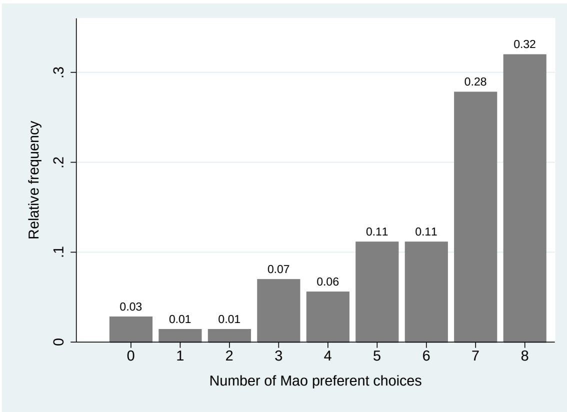
Figure 1.5: Distribution of the number of prudent choices by subjects



choices is equal to 8 as it would be for arbitrary choices ($p = 0.0000$, two-sided one-sample Wilcoxon signed-rank test).[19]

In stage MAO, 77.08% of all questions are answered in a Mao-preferent way. Figure 1.6

Figure 1.6: Distribution of the number of Mao-preferent choices by subjects



illustrates the relative frequencies of subjects' Mao-preferent choices. Each subject has been, on average, Mao-preferent in 6.16 out of 8 questions with a standard deviation of 2.01. The median (mode) of Mao-preferent choices is 7 (8). Also, this behavior differs significantly from arbitrary choices ($p = 0.0000$, one-sample Wilcoxon signed-rank test).

[19]In the following, all statistical tests are *two-sided* if not indicated differently. The Wilcoxon signed-rank procedure assumes that, under the null hypothesis, the sample (of frequencies per individual) is randomly taken from a population, with a symmetric frequency distribution. The symmetric assumption does not assume normality. As an alternative, a two-sided one-sample sign-test, implying the same null and alternative hypothesis (but without the symmetry assumption), would also lead us to reject the null ($p = 0.0004$).

1.5.2 Within subject analysis

Our preliminary analysis suggests substantial evidence for prudent and Mao-preferent behavior. This subsection is concerned with their relationship at an individual level (Research Question 1). For starters, we observe a significant positive correlation (which is a symmetric measure of association) of $\rho = 0.2844$ between prudent and Mao-preferent choices ($p = 0.0155$, Spearman rank correlation test).

We now show that the actual relationship is asymmetric. To this end, we categorize subjects' responses in stages ES and MAO according to the frequency of prudent and Mao-preferent choices, respectively. These categorizations are somewhat arbitrary. However, the qualitative conclusions stay the same when changing the categorizations by plus or minus one. Subjects who answered 12 or more (4 or less) out of 16 questions prudently are said to be *prudent* (*imprudent*). Those subjects who answered 5 to 11 questions prudently are classified as *indifferent*. Similarly, subjects are classified as *Mao-preferent* (*not Mao-preferent*) if they have answered 7 or 8 (0 or 1) out of 8 questions in favor of the lottery with the positive (negative) skewness. When answering 2 to 6 questions in a Mao preferent manner, subjects are allotted to the category *indifferent*.

Table 1.3 cross-tabulates the absolute frequencies of subjects according to the categories. Let us first analyze prudence and the Mao preference separately. 34 (47.22%) of all 72 subjects are prudent whereas only 6 (8.33%) are imprudent. Note again that this gives a very different picture, as compared to looking at the aggregate responses only. Deck and Schlesinger (2010) report that very few subjects always decided imprudently (2%) and only 14% were always prudent in their six decision tasks. The Mao preference is more widely observed than prudence, as 43 (59.72%) of all subjects exhibit it, whereas only 3 (4.17%) do not.[20] This complies with our arguments made in Sections 1.2, as it shows that, also empirically, the Mao preference is a weaker preference than prudence. The difference in prudent and Mao-preferent observations immediately indicates that Mao lotteries are

Table 1.3: Contingency table on categories

	Not Mao-preferent	Indifferent	Mao-preferent	Total
Imprudent	0	3	3	6
Indifferent	2	13	17	32
Prudent	1	10	23	34
Total	3	26	43	72

[20]Tarazona-Gomez (2004) finds 63% of the subjects to be 'prudent' under the assumption of third-order moment preferences.

indeed not suitable to test for prudence.

The conditional frequency Prob(Mao-preferent |prudent) that a prudent individual exhibits the Mao preference is 67.65%, whereas Prob(not Mao-preferent |prudent) is only 2.94%.[21] The chance for a prudent individual to be Mao-preferent is thus about 23 times higher than not being Mao-preferent. The analysis of the reverse statement does not provide such a clear-cut picture. The conditional frequency Prob(prudent |Mao-preferent) is given by 53.49% whereas Prob(imprudent |Mao-preferent) equals 6.98%. Thus, the chance of being prudent given that an individual is Mao-preferent is about 8 times higher for an individual that is not Mao-preferent. This finding, however, is not very reliable as there are only 3 subjects who were not Mao-preferent.

In short, we see that knowing about an individual's preference towards the Mao lotteries gives some information about whether the individual is prudent. This finding also hints in the 'right' direction, as being Mao-preferent increases the probability of being prudent. It suggests that most prudent individuals exhibit the Mao preference, whereas Mao-preferent individuals may not be prudent. This implies that the Mao preference may not be sufficient to make conclusions whether an individual is prudent. Thus, there seems to be more to prudence than skewness seeking. In particular, it would lead us to reject third-order moment preferences. The finding can also be interpreted as a robustness check for our method to test for prudence. Those subjects it diagnoses as prudent, consistently with theory, are Mao-preferent.

1.5.3 Influences on prudent behavior

We now investigate what types of ES questions are more likely to be answered prudently. In general, we find that the particular choice of the prudence lottery pair has a strong impact on the 72 subjects' decisions. Relative frequencies range from 50.00% to 75.00% with a standard deviation of 8.11% from the reported mean of 65.10%. This is evidenced by the relative frequencies shown in the last column of Table 1.1 in Subsection 1.4.1 which also shows the factor levels for each question. In order to determine what particular elements in the definition of prudence cause these differences we investigate Factors A, B, C and D according to Research Questions 2 to 5.

As formulated in Research Question 2 we are interested whether the fixed amount k being a gain or a loss (Factor A) influences subjects' decisions. When k is a loss, 66.32%

[21] If we exclude subjects who were indifferent at least at one stage these numbers become 95.6% and 4.2%, respectively.

of responses are prudent, whereas slightly less responses are prudent (63.89%) when k is a gain. Test statistics of a Fisher-Pitman permutation test for paired replicates in Table 1.4 show that this difference is insignificant ($p = 0.5253$ and $p = 0.5008$, respectively). This implicitly suggests that 0 as a focal point did not influence subjects' behavior.

Considering Factor B, 64.76% of choices are prudent if the wealth level x is high ($x_1 = 160$) and 65.45% of choices are prudent if x is low ($x_2 = 80$) what indicates an insignificant difference (see the test results in Table 1.4).

Research Question 4 asks whether a framing of the decision task (Factor C) influences subjects' decisions. The level of Factor C influences prudent choices substantially, as 67.36% of the choices are prudent if the level is "add ϵ" and 62.85% if the level is "add $-k$". Test statistics show that differences are significant below a 10% level.

Result 1. *Framing of the decision task influences subjects' decisions. Weakly significant more subjects answer questions prudently if the zero-mean risk (ϵ) has to be added to the 50/50 gamble compared to the fixed amount (k).*

In essence, Result 1 shows that the decision task involving subjects' conscious consideration about another risky event leads to more prudent choices, whereas when asked to add a fixed amount subjects make slightly more imprudent choices. When looking at the interaction of Factors A and C weakly significantly more choices are prudent whenever i) the fixed amount is a loss ($k_1 = 40$) and subjects are asked to "add ϵ" and ii) the fixed amount is a gain ($k_2 = -40$) and they are asked to "add $-k$" ($p = 0.0690$).

In short, we find that subjects' decisions are neither significantly influenced by the fixed amount being a gain or a loss nor by the wealth level. These findings are in line with behavioral patterns reported by Deck and Schlesinger (2010). They report that the relative size of the zero-mean risk is not influential. In contrast to their findings, our behavioral data evidence that framing of the decision task weakly influences subjects' choices.

Factor D considered in Research Question 5 is most significant (see Table 1.4). At its low level (negative kurtosis difference), 68.58% of subjects' choices are prudent. If Factor D is at its high level (positive kurtosis difference), 61.63% of the choices are prudent. For questions ES9 (largest positive kurtosis difference in the experiment) and ES10 (largest negative kurtosis difference, other factors like in ES9), 50.00% and 75.00% of answers are prudent, respectively. Note again that for the prudence lotteries a negative kurtosis

Table 1.4: Analysis of prudent choices for different factor levels

Factor	Level	Relative frequency of prudent choices	p
A	$k_1 = 40$	0.6632	0.5008
	$k_2 = -40$	0.6389	
B	$x_1 = 160$	0.6476	0.8362
	$x_2 = 80$	0.6545	
C	add $-k$	0.6285	0.0677
	add ϵ	0.6736	
D	$\kappa(B_3) - \kappa(A_3) < 0$	0.6858	0.0121
	$\kappa(B_3) - \kappa(A_3) > 0$	0.6163	

This table shows relative frequencies of prudent choices for each factor level. p values of a two-sided Fisher-Pitman permutation test are shown testing for differences between relative frequencies of prudent choices for given levels of each factor.

difference is equivalent to ϵ being left-skewed, i.e., the ballot box displayed on subjects' screens contains more white balls (implying a small gain) than yellow balls (implying a high loss).

Result 2. *The particular choice of the zero-mean risk ϵ strongly influences subjects' decisions. Significantly more subjects decide prudently if ϵ is left-skewed.*

One intuition supporting Result 2 is that a prudent individual may consider a negatively skewed zero-mean risk as a "bigger" harm. Hence, there is a greater tendency for disaggregating the harms of the sure loss and the zero-mean risk. In Section 1.2, we showed that ϵ being left-skewed implies a smaller kurtosis for the prudent choice. An interpretation is that in this case the prudent choice implies a smaller likelihood of extreme events to occur. A prudent individual, however, would seek the higher skewness of the prudent lottery choice irrespectively of its kurtosis. She must not deviate from her preference if the additional risk is not too harmful to her. This was shown in the theoretical part of this paper and was referred to as the kurtosis robustness feature of prudence. Thus, Result 2 is a major finding of our experiment as it confirms its relevance empirically. It emphasizes the importance to use several lotteries to test for prudence in order to reflect the statistical diversity which is implicit in Eeckhoudt and Schlesinger's definition of prudence. As the kurtosis of the prudence lotteries matters, the significance of Factor D also shows that there is more to prudence than skewness seeking.

The above observations are generalized and detailed in Ebert (2010). He shows that a mixed risk averse decision maker[22] has higher utility from proper risk apportionment if ϵ is left-skewed. Thus our result also gives some support for mixed risk aversion.

1.5.4 Testing for moment preferences

In this section, we test directly for moment preferences. The Mao pairs were calibrated to the ES pairs according to Proposition 4 in terms of the first three moments. Lottery pair Mao 1 is calibrated to lottery pair ES1 and ES2, Mao 2 is calibrated to ES3 and ES4, and so on. We investigate whether there is a stronger association between subjects' decisions over such corresponding lottery pairs than to those over the remaining ones.

For each ES question—paired with any Mao question—we set up 8 contingency tables. That equals 128 2×2-contingency tables, in total, among which are 16 tables for corresponding Mao and ES pairs. As a measure of association we use the *phi coefficient* (r_ϕ). Each contingency table comprises the four categories i) prudent, Mao-preferent, ii) prudent, not Mao-preferent, iii) imprudent, Mao-preferent and iv) imprudent, not Mao-preferent.

The results shown in Table 1.5 are that, for 7 out of 16 comparisons, the degree

Table 1.5: Correlation (r_ϕ) between Mao and ES pairs

	MAO1	MAO2	MAO3	MAO4	MAO5	MAO6	MAO7	MAO8	p-value
ES1	**0.1491**	0.0087	0.0016	0.0888	0.0262	0.0940	0.1027	0.0170	0.0078
ES2	**0.0028**	0.0194	0.0007	0.0032	0.0045	0.0036	0.0032	0.0005	0.7031
ES3	0.2130	**-0.0510**	0.0400	-0.0041	0.0117	0.2050	0.2273	-0.1544	0.9688
ES4	0.1050	**0.0410**	0.1480	0.0960	-0.1035	0.1478	-0.1409	0.1011	0.5859
ES5	0.1315	0.0127	**0.0974**	-0.0174	-0.0029	0.1229	0.0769	0.0223	0.0547
ES6	0.2800	0.1888	**0.1312**	0.0806	-0.1176	-0.0075	0.1282	0.0133	0.1953
ES7	0.1272	0.0555	0.1654	**0.1120**	-0.0127	0.1692	0.1663	0.0508	0.3750
ES8	0.0836	0.1888	0.1312	**0.0022**	-0.1176	-0.0748	0.1282	0.1501	0.9297
ES9	0.1206	0.1491	-0.0702	0.0361	**0.0342**	0.0000	0.0327	-0.0945	0.3672
ES10	0.0348	0.0861	0.2026	0.0625	**0.0987**	0.0000	0.1322	0.0910	0.3125
ES11	0.1713	0.0385	0.0544	-0.0653	-0.2561	**0.0320**	-0.0253	-0.1383	0.1406
ES12	-0.0940	0.0410	0.1480	0.0166	0.1223	**-0.0569**	0.0030	0.0318	0.9922
ES13	0.2597	0.3107	0.2243	0.2659	0.2537	0.1413	**0.3099**	0.1112	0.0156
ES14	-0.1127	0.0270	-0.0212	0.0022	0.1050	-0.2094	**0.1992**	-0.0551	0.0078
ES15	0.1315	0.0887	0.1689	0.0562	0.0668	-0.0035	0.1435	**-0.2989**	1.0000
ES16	-0.0075	0.1387	0.1654	0.1120	0.1400	0.1000	0.0933	**0.0508**	0.9766

The values on the diagonal in bold face indicate r_ϕ for the corresponding ES and Mao pairs. p-values are shown for a one-sided Fisher-Pitman permutation test for paired replicates.

of association between the Mao and the corresponding ES pair is stronger compared to

[22] An expected utility maximizer is called mixed risk averse if she is nth-degree risk averse of all orders, i.e., $sgn(u^{(n)}) = (-1)^{n+1}$ for all n. All of the commonly used utility functions exhibit mixed risk aversion. In particular, such a decision maker is prudent.

the remaining ones. For 4 out of these 7 associations the difference is weakly significant at a 6% level as indicated by test results of a one-sided Fisher-Pitman permutation test (see last column of Table 1.5). The probability that the degree of association of a corresponding lottery pair is largest by coincidence is one out of eight. For 7 successes out of 16 observations, the null hypothesis (that the probability of a success on a single trial is 1/8) has to be rejected ($p = 0.0019$ two-sided binomial-test).

Result 3. *For a significant number of ES pairs, the number of prudent choices is 'closest' to the number of Mao-preferent choices for the corresponding Mao pair. This indicates that the first three statistical moments have some predictive power for prudence.*

The weak correlation between moments and preferences also supports the necessity of appropriate lottery calibration. This way it can be ruled out that measured effects are only due to different parameter ranges among lotteries. Further, the results are in line with theoretical findings of Brockett and Garven (1998) that subjects' decisions in the experiment cannot be explained completely by the first three moments only. In particular, prudence is not perfectly captured by skewness seeking.

1.5.5 Risk aversion and individual characteristics

According to Eeckhoudt and Schlesinger (2006), prudent individuals can be risk averse, risk neutral or risk loving. This is confirmed by our data where a substantial proportion of subjects is risk averse (87.50%). Among the prudent (non-prudent, i.e., imprudent and indifferent) subjects 1 (3) are risk-neutral, 29 (34) are risk averse and 3 (1) are risk loving. The degree of risk aversion of prudent and non-prudent individuals does not differ significantly ($p = 0.8106$, Mann-Whitney U-test).

The age of the 72 participants is, on average, 24.25 years; the youngest individual is 19 the oldest 42 years of age. 41 were female and 31 were male. According to Mann-Whitney U-tests, neither age nor gender had a significant influence on the number of prudent answers observed in our experiment.

1.6 Conclusion

Currently, the share between theoretical and empirical literature on prudence is very unbalanced. Numerous behavioral implications of prudence have been pointed out, but there is very little empirical, i.e., experimental, research on prudence to support the relevance and validity of these theories. To get there, in this study we propose, implement and check for robustness a method testing for prudence in a laboratory setting.

We constructed a set of 16 prudence lottery pairs (Eeckhoudt and Schlesinger 2006) that not only reflect skewness seeking, but also the kurtosis robustness feature of prudence. As shown, the latter is also a characteristic feature of prudence. Its origin lies in the skewness of the zero-mean risks and we show how to implement such risks in the experiment. To this end, we propose a new ballot box representation of compound lotteries for application in experiments. It is very easy to understand and translates naturally from subjects' decision screens to the real-world draw of the lotteries.

In the experiment, indeed, the choice of the zero-mean risk significantly affects subjects' decisions. Thus, we find that prudence does not boil down to skewness seeking. Prudence is observed on the aggregate as well as at the individual level. 65% of responses are prudent and we classify 47% of individuals as prudent and 8% as imprudent. The number of prudent responses varies substantially from 50% to 75% for different prudence lottery pairs. This should be taken into account when testing for prudence.

Moreover, this study contains a lottery calibration theorem that allows the researcher to construct binary lotteries with desired first three moments. We illustrate how this theorem can be used to construct lotteries in the desired parameter range. We also present a statistical characterization of the lotteries of Mao (1970) and show that preference between such lotteries is purely determined by the difference in their skewness.

Given the observed presence of prudence, further experimental research could focus on the empirical validation of prudent behavior. For example, the probably most famous prediction that prudent people exhibit larger precautionary saving has received little attention yet. Moreover, the proposed method could be easily adapted to test for temperance and associated theories.

Chapter 2

Joint measurement of risk aversion, prudence and temperance

2.1 Introduction

The concept of *risk aversion* plays a key role in analyzing decision-making under uncertainty. An established characterization is that an individual preferring a payoff with certainty over a risky payoff with the same mean is said to be risk averse (Gollier 2001, p.18). Alternatively, Rothschild and Stiglitz (1970) state that a risk averse individual dislikes any mean-preserving spread of the wealth distribution. Within an expected utility (EU) setting, these two characterizations coincide and are equivalent to the utility function being concave.

Unlike the term risk aversion suggests, it is not a concept to describe an individual's risk preferences exhaustively. It is just one piece in the puzzle, complemented by higher-order risk preferences. *Prudence* (third-degree risk aversion) and *temperance* (fourth-degree risk aversion) are lesser known traits affecting behavior towards risk. Although Kimball (1990) coined the term 'prudence', its implications have been used in assessing a precautionary demand for saving much earlier by Leland (1968) and Sandmo (1970). In particular, they show within an EU setting how a risky future income does not guarantee that a consumer increases saving unless the individual exhibited prudence. The notion of 'temperance' was also introduced by Kimball (1992). Temperance refers to the fact that the advent of an unavoidable risk should lead an individual to reduce the exposure to another risk even if the two risks are statistically independent.

Recently a large theoretical literature on the implications of higher-order risk preferences has emerged. Eeckhoudt and Gollier (2005) analyze the impact of prudence on prevention, i.e., the action undertaken to reduce the probability of an adverse effect to occur. This also plays an important role in a medical decision-making context (see Courbagé and Rey 2006). Esö and White (2004) show that there can be precautionary bidding in auctions when the value of the object is uncertain and when bidders are prudent. Likewise, White (2008) analyzes prudence in bargaining. Treich (forthcoming) shows that prudence can decrease rent-seeking efforts in a symmetric contest model. Eeckhoudt and Schlesinger (2008) show that temperance is necessary and sufficient for an increase in downside risk of future labor income to always increase the level of precautionary savings. Other examples are insurance demand (e.g., Fei and Schlesinger 2008) or life-cycle investment behavior (e.g., Gomes and Michaelides 2005). By necessity this is not a complete list of applications.

Prudence and temperance also play key roles in aversion to negative skewness and kurtosis, respectively. Prudence has been shown to be equivalent to aversion to increases in downside risk as defined by Menezes et al. (1980). A downside risk increase does not change mean and variance of a prospect, but does decrease its skewness. Likewise, an increase in outer risk increases kurtosis but leaves the first three moments of a distribution unchanged. Menezes and Wang (2005) show that temperance is equivalent to outer risk aversion.

More recently, both prudence and temperance have been characterized outside an EU context by Eeckhoudt and Schlesinger (2006) as preferences over 50/50 lottery pairs. Their definition, which is shown to be equivalent to the ones mentioned, is particular appealing for experimental purposes. Prudence is defined as a preference for disaggregating a zero-mean risk and a sure reduction in wealth across two equally likely states of nature. Analogously, temperance is a preference for the disaggregation of two independent zero-mean risks.

On the empirical side, there is an extensive literature on the measurement of risk aversion in numerous empirical settings (e.g., Barsky et al. 1997) as well as in various experiments. Focusing on experiments, almost as large as the number of experimental studies is the diversity in procedures. Two well established methods based on binary lottery choices are the multiple price list (e.g., Schubert et al. 1999, Holt and Laury 2002, Barr and Packard 2002) and random lottery pairs technique (e.g., Grether and Plott 1979, Hey and Orme 1994). An alternative approach comprises a selection task from an ordered set of lotteries (e.g., Binswanger 1980, Eckel and Grossman 2008 a). Another prominent method is the Becker-DeGroot-Marschak auction where a certainty equivalent is elicited (Becker et al. 1964, Harrison 1986, Loomes 1988). Related to the latter method Wakker and Deneffe (1996) propose a certainty equivalent technique with endogenous probabilities. In Dohmen et al. (forthcoming a) subjects decide between safe and risky options in a variant of the so-called switch multiple price list technique.[1] Another experimental method based on the proper risk apportionment model will be proposed in this paper.

In contrast, there are few empirical studies on higher-order risk attitudes. Dynan (1993), Carrol (1994) and Carroll and Kimball (2008) trace prudence via the precautionary savings motive. We are not aware of an empirical study testing for temperance.

Laboratory experiments could be used to investigate prudence and temperance as well as the associated theories and behavioral traits in a more controlled environment. Research in this direction has just started. The first attempt in this direction was made by Tarazona-Gomez (2004) who finds weak evidence for prudence. Her experiment relies on a certainty equivalent approach involving lotteries with several different outcomes. It is based on strong assumptions within expected utility theory, in particular, a truncation of the utility function. The only other papers testing for prudence are Deck and Schlesinger (2010) and Ebert and Wiesen (2009) which test for prudence using the lotteries of Eeckhoudt and Schlesinger (2006). Both papers find significant support for prudence (61% and 65% of responses, respectively). Ebert and Wiesen (2009) motivate and show that the choice of the zero-mean risk considered in Eeckhoudt and Schlesinger's proper risk apportionment model significantly influences subjects' decisions. Deck and Schlesinger (2010) also test for temperance and find weak evidence for intemperate behavior.

These studies *test for the direction* of third- or fourth order risk preferences, but do not measure their *intensity*. Subjects make several lottery choices and inference is made on the count of prudent (temperate) choices. In particular, such a design makes it difficult to compare the relative importance of prudence or temperance for a given individual.

Thus the aim of the present paper is, firstly, to present a method to measure the *intensity* of risk aversion, prudence and temperance. Secondly, this is done jointly such that we can compare their relative importance to a given individual. The approach is not based on EU. We define higher-order risk premia á la Arrow-Pratt within the proper risk apportionment model of Eeckhoudt and Schlesinger. More specifically, we measure the smallest amount that must be added to the lottery with more 2nd- (3rd-, 4th-) degree risk that makes an individual prefer this lottery over the one with less 2nd- (3rd-, 4th-)

[1]See Harrison and Rutström (2008) for a comprehensive overview on different experimental methods to elicit risk aversion.

degree risk. This implies a clean tradeoff between nth degree risk and expected wealth. The lotteries in the experiment are calibrated such that premia for different degrees of risk are comparable. We also show how these premia are related to indices of higher-order risk attitudes defined in the literature just recently (Jindapon and Neilson 2007 and Denuit and Eeckhoudt forthcoming b).

Our experimental method is a combination of the compound lottery display introduced in Ebert and Wiesen (2009) and a multiple price list technique which is popular from the Holt and Laury (2002) experiment. A within-subject design is applied to measure risk premia of different orders for all subjects. This design in turn is embedded in a between-subject factorial design used to test our approach for robustness to typical manipulations of the experimental setup.

We find substantial evidence for risk aversion, prudence and temperance. Most interestingly, subjects demand a significantly higher downside risk premium compared to the 2nd-order risk premium. This highlights the importance of prudence and likewise questions the excessive focus on risk aversion in the economics literature, both theoretical and empirical. In particular, the literature contains numerous different experimental methods to measure risk aversion, but this paper constitutes the first approach to measure prudence. The outer risk premium is smallest and significantly different from both the downside and outer risk premia. However, it is still significantly positive which indicates that most subjects are temperate, contrary to the tendency observed in Deck and Schlesinger (2010). For a given subject, we find a positive correlation between premia demanded, in particular for prudence and temperance. Thus, given the assumption of EU, our experiment supports the assumption of mixed risk aversion (Caballe and Pomansky 1996) which is exhibited by all the commonly used utility functions (Brockett and Golden 1987).

Moreover, we controlled the number of male and female participants in order to check for possible gender differences. Differences between women and men in risk attitudes are well documented in the experimental economics literature. Most evidence suggests that women perceive risks as greater, engage in less risky behavior, and choose alternatives that involve less risk, see Eckel and Grossman (2008 b) and Croson and Gneezy (2009) for reviews of the relevant literature. We show that this is also the case for higher order risk attitudes. Women are significantly more risk averse, more prudent and more temperate than men.

The remainder of this paper is as follows. In Section 2.2 we review the proper risk apportionment model and define risk premia of higher-orders. In Section 2.3 we explain our experimental approach to elicit these premia. In Section 2.4 we give the results of the experiment and in Section 2.5 we conclude.

2.2 Proper risk apportionment approach to elicit higher-order risk premia

Within the expected utility (EU) framework, assuming differentiability of a utility function u, risk aversion, prudence and temperance are defined as $u'' < 0$, $u''' > 0$ and $u^{(4)} < 0$, respectively. However, our experimental methodology is not based on EU but on the proper risk apportionment model of Eeckhoudt and Schlesinger (2006). Therefore, risk aversion, prudence and temperance are defined as a preference over lottery pairs.

We first define (2nd-degree) risk aversion. Let x be the individual's wealth and $k, r > 0$ are fixed monetary amounts. With $B_2 = [x - r, x - k]$, for example, we denote the 50/50 gamble B_2 that has equally likely payoffs $x - r$ and $x - k$. An individual is called risk averse if she prefers B_2 to $A_2 = [x - r - k, x]$ for arbitrary parameter values x, r and k. The lotteries are displayed in Figure 2.1. Thus, a risk averse individual prefers to disaggre-

Figure 2.1: Risk aversion lottery pair (A_2, B_2)

This figure shows lotteries of the type used to measure risk aversion in the experiment. x is the subject's endowment and $-r$ and $-k$ denote sure reductions in wealth. To imagine a prudence lottery pair (A_3, B_3), simply replace the $-r$ with a zero-mean risk ϵ_1. To imagine a temperance lottery pair (A_4, B_4), additionally replace $-k$ with a second independent zero-mean risk ϵ_2.

gate unavoidable losses $-r$ and $-k$ across states of nature. This preference is equivalent under EU to $u'' < 0$, as shown in Appendix A.2.1.[2] The preference is also equivalent to a preference for decreases in risk in the sense of Rothschild and Stiglitz (1970).

In order to define prudence (3rd-degree risk aversion, downside risk aversion), the sure reduction in wealth $-r$ in the definition for risk aversion (also illustrated in Figure 2.1) is replaced with a zero-mean risk ϵ. That is, an individual is called prudent if she prefers

[2] Eeckhoudt and Schlesinger (2006) originally considered the lotteries $\tilde{B}_2 \equiv 0$ and $\tilde{A}_2 = [0, \epsilon]$, where ϵ is a zero-mean risk, and show that preferring $\tilde{B}_2$ over $\tilde{A}_2$ for all ϵ is equivalent to $u'' < 0$. We use the lotteries B_2 and A_2 instead because a certainty effect could distort experimental results when using $\tilde{B}_2$ and $\tilde{A}_2$.

$B_3 = [x-k, x+\epsilon]$ over $A_3 = [x, x-k+\epsilon]$ for all wealth levels x, sure wealth reductions $-k$ and zero-mean risks ϵ. That is, a prudent individual prefers to disaggregate an unavoidable risk and a loss across different states of nature. Equivalently, an unavoidable risk is preferred when wealth is higher. Eeckhoudt and Schlesinger (2006) show that this preference is equivalent to $u''' > 0$ in EU or to a preference for decreases in downside risk as defined by Menezes et al. (1980).

Finally, temperance (4th-degree risk aversion, outer risk aversion) is defined as a preference of $B_4 = [x+\epsilon_1, x+\epsilon_2]$ over $A_4 = [x, x+\epsilon_1+\epsilon_2]$ where ϵ_1 and ϵ_2 are two independent zero-mean risks. Under EU, this preference is equivalent to $u^{(4)} < 0$ and it is also equivalent to a preference for decreases in outer risk as defined by Menezes and Wang (2005).

Eeckhoudt and Schlesinger (2006) define a nesting process to construct lotteries B_n and A_n from the lotteries B_{n-2} and A_{n-2}. Then they show that the preference B_n over A_n is equivalent to $(-1)^{(n)}u^{(n)} < 0$ under EU which was labeled nth-degree risk aversion by Ekern (1980). An individual might be, for example, risk-loving and prudent (imprudent) just as it might be risk averse and prudent (imprudent). If an individual prefers B_n over A_n for all n she is called *mixed risk averse*, see Caballe and Pomansky (1996). Under EU this means that her utility function is increasing with derivatives of alternating sign. It is interesting to note that all the commonly used utility functions imply mixed risk aversion, see Brockett and Golden (1987). Ebert (2010) showed that the utility of downside risk diversification is maximal if and only if the EU decision maker is mixed risk averse. By measuring three different degrees of risk aversion, we obtain a testable hypothesis for mixed risk aversion.

Our experiment aims to elicit risk premia for nth-degree risk aversion for $n = 2, 3, 4,,$ i.e., a (2nd-degree) risk premium m^{RA}, a downside risk (imprudence) premium m^{PR} and an outer risk (intemperance) premium m^{TE}. For example, in the case of risk aversion, for every individual we aim to elicit m^{RA} where she is indifferent between $B_2 = [x-r, x-k]$ and $A_2 + m^{\text{RA}} = [x + m^{\text{RA}}, x-r-k+m^{\text{RA}}]$. For a (2nd-degree) risk-loving individual, m^{RA} will be negative. Unlike in the experiments of Deck and Schlesinger (2010) and Ebert and Wiesen (2009) we, thus, obtain a measure of the intensity of the risk attitude rather than only a test of preference direction.

Before we explain the procedure to implement our higher-order risk premia approach to measure higher-order risk preferences in a laboratory experiment, it is interesting to relate it to the very recent theoretical literature on higher-order intensity measures. Generally, this literature is concerned with generalizing the measures introduced in Arrow (1965), Pratt (1964) and Ross (1981) to higher-orders. However, keep in mind that our approach is not based on EU and, in particular, does not rely on any of the assumptions or approximations frequently observed in that literature.

Kimball (1990, 1992) established a link between $\frac{u'''}{u''}$ and the intensity of precautionary savings. Chiu (2005) gave a choice-theoretic foundation of this measure paralleling that of Arrow and Pratt being generalized to nth order by Denuit and Eeckhoudt (forthcoming a). Modica and Scarsini (2005) show that $\frac{u'''}{u'}$ is a natural extension of the Ross measure of risk aversion and suggest that it is also locally a good measure for the intensity of downside risk aversion. In particular, the difference in risk premia of random variables with equal mean and variance can approximately be written as the product of $\frac{u'''}{u'}$ and the difference in their third moments. This holds locally at any wealth level x. Jindapon and Neilson (2007) and Denuit and Eeckhoudt (forthcoming b) generalize their results and conclude that $(-1)^{n+1}\frac{u^{(n)}}{u'}$ is also locally an appropriate index of nth order risk attitude. For example, $-\frac{u^{(4)}}{u'}$ is an appropriate measure for kurtosis aversion. Crainich and Eeckhoudt (2008) more specifically consider a premium for Eeckhoudt and Schlesinger's prudence lotteries and relate it to $\frac{u'''}{u'}$. In derivations similar to theirs, in Appendix A.2.1 we show that for the lotteries in our experiment we have

$$-\frac{u''(x)}{u'(x)} \approx \frac{2m^{\mathrm{RA}}}{r(k-m^{\mathrm{RA}})} \tag{2.1}$$

$$\frac{u'''(x)}{u'(x)} \approx \frac{4m^{\mathrm{PR}}}{\sigma^2(k-m^{\mathrm{PR}})} \tag{2.2}$$

$$-\frac{u^{(4)}(x)}{u'(x)} \approx \frac{8m^{\mathrm{TE}}}{\sigma_1^2\sigma_2^2} \tag{2.3}$$

where $\sigma^2 = E[\epsilon^2]$, $\sigma_1^2 = E[\epsilon_1^2]$ and $\sigma_2^2 = E[\epsilon_2^2]$ denote the variances of the zero-mean risks.

Each intensity measure is increasing in the corresponding premium $m^{\bullet}$. If we further assume that $m^{\mathrm{RA}}r$ and $m^{\mathrm{PR}}\sigma^2$ are small compared to rk and $\sigma^2 k$, respectively, we can add some more interpretation. The difference in variance of the risk aversion lotteries is $-rk$, the difference in the unstandardized central third moment of the prudence lotteries is $0.75\sigma^2 k$ and the difference in the unstandardized central fourth moment is $-1.5\sigma_1^2\sigma_2^2$.[3] Thus in this case, each premium of order n we measure is proportional to the corresponding intensity measure of order n and to the difference in moments of order n. While we think

[3]Because of centralization this holds for the lotteries with or without premium. See, e.g., Ebert (2010) for the computations.

that these results provide valuable intuition it should be noted that the approximations of the above formulas are rather strong. Again, our experimental procedure and results do not rely on these approximations, and not even on EU theory at all.

2.3 Experimental design and procedure

In this section we first present the general set up of the experiment. Then we describe the decision situation in-depth. We further discuss our experimental methodology and the choice of parameters. Finally, we describe the experimental procedure.

2.3.1 General design

In our experiment, we present subjects with a menu of pairwise lottery choices that permits us to identify subjects' degree of risk aversion, prudence and temperance. Thereby we extend the methodology of Deck and Schlesinger (2010) and Ebert and Wiesen (2009) who test for higher-order risk attitudes in a yes-or-no fashion. Further, because we use a within-subject design, we can compare the intensity of risk attitudes of orders 2,3 and 4 at an individual level. That is, we can investigate their relative importance to subjects. The main idea of the method is to combine a multiple price list format[4] with the ballot box representation of compound lotteries introduced in Ebert and Wiesen (2009).

Overall subjects make 120 decisions. After the experiment one decision is randomly selected to determine that subject's payoff.[5] The 120 decisions are divided into 20 decisions on each out of 6 different decision screens. One screen is for risk aversion (stage RA), three screens are for prudence (stage PR, tasks PR1, PR2, PR3) and two screens are for temperance (stage TE, tasks TE1, TE2). On each screen, subjects make 20 choices over a lottery pair as introduced in Section 2.2, where each decision is for a different value of the risk premium. For example, in stage RA subjects decide between B_2 and $A_2 + m^{\mathrm{RA}}$

[4]Besides the studies mentioned in the Introduction prominent examples of studies employing a multiple price list method to elicit risk attitudes are Murnighan et al. (1988) and Gonzalez and Wu (1999).

[5]It has become quite common in economic experiments to elicit a series of choices from participants and then to pay for only one randomly selected choice; see Baltussen et al. (2010) for a fine overview. The random choice payment technique enables the researcher to observe a large number of individual decisions for a given research budget. However, the important question arises whether subjects behave as if each of these choices involves the stated payoffs. This issue has been analyzed, among various other setups, in experiments with pairwise lottery choice problems similar to our experiment. For example, Starmer and Sugden (1991) found clear evidence that under random payment subjects isolate choices as if paid for each task. Similar evidence was reported by Beattie and Loomes (1997) and Cubitt et al. (1998). In a lottery experiment with a multiple price list format Laury (2005) reports no significant difference in choices between paying for 1 or all 10 decision.

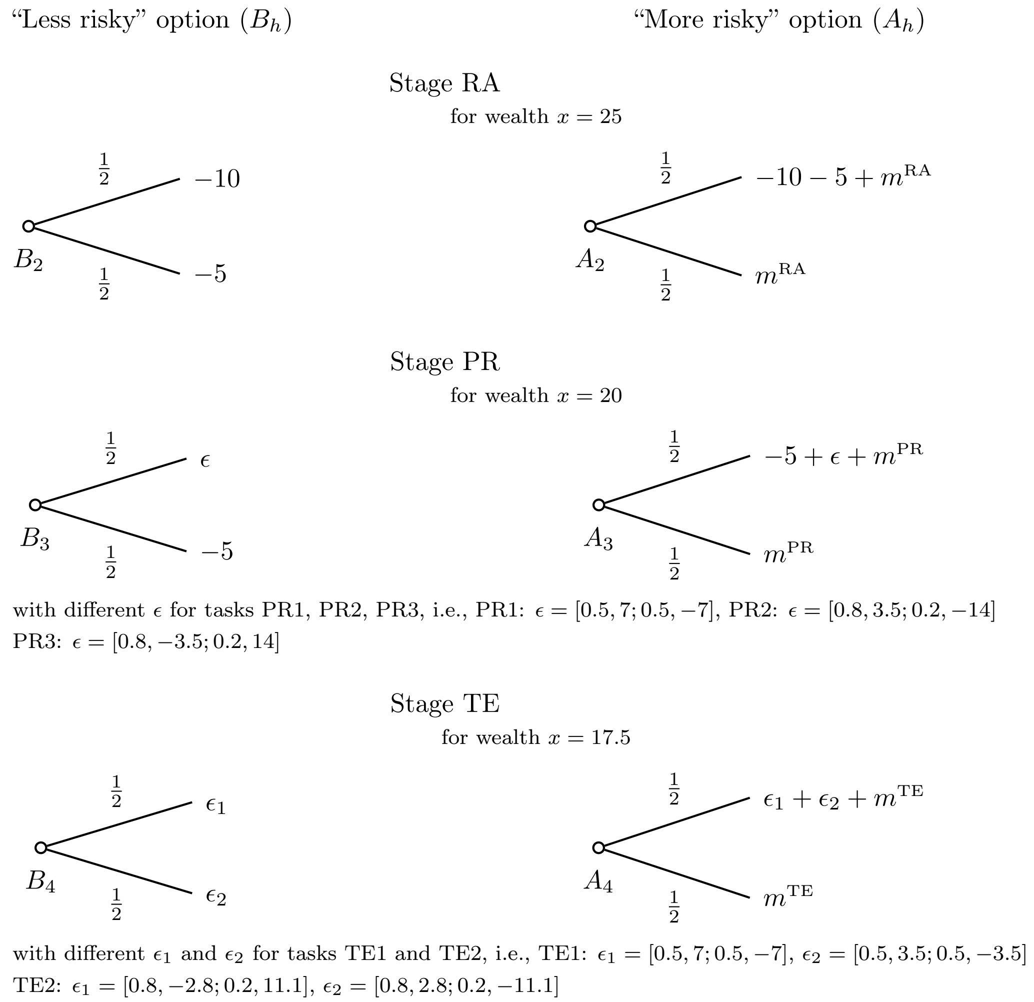
for $k = 5$ and $r = 10$ where m^{RA} takes the 20 values in EUR[6]

$$-2.50, -2.25, \ldots, -0.25, 0.00, 0.25, \ldots, 2.00, 2.25.$$

The values for m^{PR} and m^{TE} follow the same grid within one experimental session.[7] Figure 2.2 schematically illustrates the lottery pairs used in the stages of the experiment.

Because outcomes of lotteries could be negative subjects received an endowment per

Figure 2.2: Lottery pairs in the stages of the experiment



This figure illustrates the lotteries used in the experiment including the risk premia. For risk aversion we measure on premium m^{RA}, for prudence temperance we measure three premia $m^{\text{PR1}}, m^{\text{PR2}}$ and m^{PR3} and for temperance we measure two premia m^{TE1} and m^{TE2}.

decision. Endowments vary across stages being 25.00 EUR in stage RA, 20.00 EUR in stage

[6]Notice that all monetary values in the experiment are indicated in EUR.

[7]In Subsection 2.3.4 we explain how we test for robustness to variations of the risk premia grid and to sequencing effects.

PR and 17.50 EUR in stage TE. The endowment is shown on subjects decision screens and, additionally, subjects are handed coupons illustrating the endowment right before each stage is started.

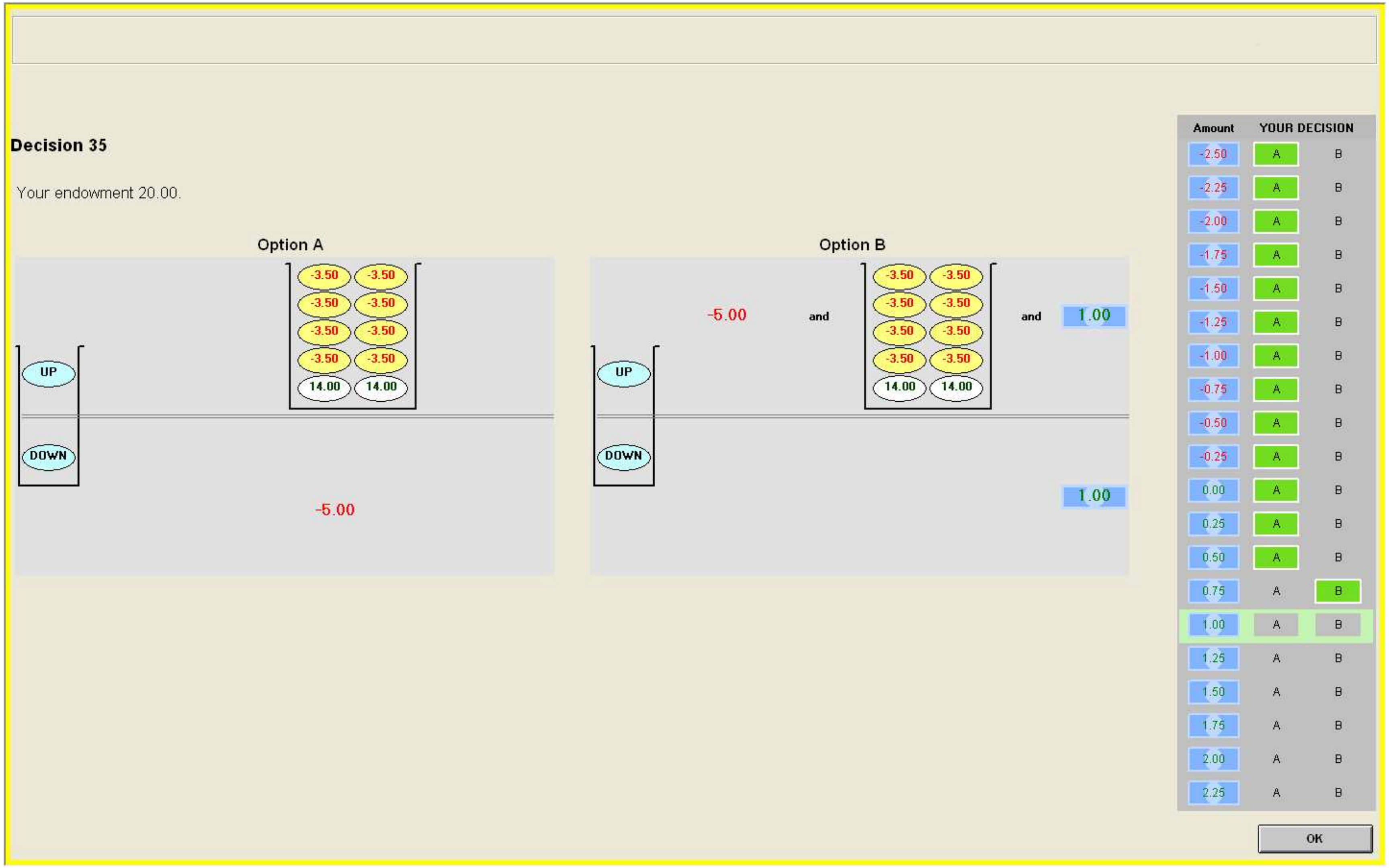
2.3.2 Decision screens

We use a computerized experiment programmed with z-Tree (Fischbacher 2007) to make use of appropriate randomization techniques explained later. While it is somewhat cumbersome to explain the decision screen in writing, it is conveniently explained to subjects in a presentation prior to the experiment. In the following we will describe the decision situation in the experimental stages in more detail. We begin with stage PR and afterwards contrast it with stages RA and TE.

An example of a decision task in stage PR (task PR3) is given in Figure 2.3. It must be understood as follows: On the upper left the number of the current decision is displayed ('Decision 35'). Below follows a statement indicating subjects' endowment which is constant for all decisions in that stage. By clicking the "OK" button on the lower right corner the subject can leave the decision screen if all 20 decisions have been made. Otherwise the subject is reminded through a pop-up window that she first has to complete all 20 decisions on that screen. The rest of the screen is divided into three panels that are displayed in darker greys than the background. From left to right, the first panel displays one possible lottery choice (Option A), the second displays the alternative lottery choice (Option B) and the far right, dark grey panel is the "decision panel."

We start with explaining the representation of the lottery displayed in the *first* panel (Option A). It consists of three items. Left is a ballot box containing two (blue) balls, labeled "Up" and "Down", respectively. Note that the panel itself is horizontally separated. The upper part contains a second ballot box that contains eight (yellow) balls labeled "-3.50" and two (white) balls labeled "14.00" and represents a gamble that yields -3.50 with probability 8/10 and 14.00 with probability 2/10. The lower part of the panel contains the entry -5.00 which represents a fixed reduction in wealth of -5.00. In total, the lottery displayed in Option A must be understood as follows. To determine its outcome, first, a ball is drawn from the ballot box containing two balls. If "Up" is realized, this means that the outcome will be determined as depicted in the upper part of the panel. That is, a draw is made from the second ballot box. Recall that the individual's endowment in this stage is 20.00. If a ball labeled -3.50 is drawn, the outcome of the lottery in Option A would be $20.00 - 3.50 = 16.50$. If a ball labeled 14.00 is drawn, the outcome would be

Figure 2.3: Sample decision screen in stage PR (task PR3)



20.00 + 14.00 = 34.00. Now suppose, in the first gamble "Down" is drawn. Then the individual faces a sure reduction in wealth of −5.00 and the outcome of the lottery in Option A would be 20.00 − 5.00 = 15.00. The ballot boxes in the screen shot aim to mimic the real world ballot boxes used to determine subjects' payoffs that are depicted in Figure 2.4.

Now consider the *second* panel in Figure 2.3, i.e., Option B. Like the first panel it is

Figure 2.4: Real world ballot boxes to determine individual payoffs

horizontally separated and contains the same two ballot boxes and the same fixed reduction in wealth −5.00. However, for Option B the −5.00 now occurs in the upper rather than in the lower part of the panel. The second difference is that both parts of the panel contain a bill labeled 1.00. To determine the outcome of the lottery in Option B, like in Option A, first a ball is drawn from the first ballot box. If "Up" is drawn, a draw is made from the second ballot box. If −3.50 is drawn, the outcome of the lottery in Option B would be 20.00 − 5.00 − 3.50 + 1.00 = 12.50. If 14.00 is drawn, the outcome of the lottery in Option B would be 20.00 − 5.00 + 14.00 + 1.00 = 30.00. If in the first draw "Down" is drawn, the outcome of the lottery in Option B would be 20.00 + 1.00 = 21.00.

Before we explain the decision panel, note that Option A depicts a prudent lottery choice of type B_3 defined in Section 2.2 and depicted in Figure 2.2. The ballot box containing the "Up" and "Down" balls represents the 50/50 gamble. The second ballot box containing ten balls represents the zero-mean risk ϵ. Likewise, Option B depicts the corresponding imprudent lottery A_3. The bill labeled 1.00 is the downside risk premium m^{PR} for the current decision. Whether the prudent or imprudent Option is displayed as "Option A", i.e., in the first panel, is randomized for every subject individually, and so is the association of payoffs with the "up" or "down"-state of the lottery.

We now explain how subjects actually indicate their decisions in the *decision* panel. By clicking one of the 20 amounts of the m^{PR}-grid that is displayed in the first column ("Amount"), the amounts m^{PR} displayed in the first panel adjust accordingly. Also the decision number in the upper left of the screen will adjust. In Figure 2.3, the subject currently has selected to make her decision for $m^{\mathrm{PR}} = 1.00$, which is why the corresponding row in the decision panel is framed in light green. To decide for Option A or Option B, respectively, she can click either "A" in the second column of the decision panel or "B" in the third column of the decision panel. The selected button then turns dark green. The subject can continue to make another decision by clicking another fixed amount in column "Amount" and click "A" or "B" in the corresponding row. In Figure 2.3, the subject chose Option A for $m^{\mathrm{PR}} = -2.50, \ldots, 0.50$, chose Option B for $m^{\mathrm{PR}} = 0.75$ and is about to make her choice for $m^{\mathrm{PR}} = 1.00$. She can also make another choice by clicking another value of m^{PR}. This way, she can also go back to a previous decision and change it. Further, the subject is free to answer the questions in a different order as suggested in the screenshot. After having made all 20 decisions, she can leave the decision screen by clicking the "OK" button. A pop-up window will appear, asking the subject to confirm or cancel. The subject is reminded that if she confirms, her decisions in this task will be final.

The decision screens for tasks in stages RA and TE are analogous to the one just explained. A lottery of type A_2 or B_2 in stage RA is displayed like A_3 or B_3 in Figure 2.3, except that the ballot box representing the zero-mean risk is replaced by the fixed amount $r = -10$. Likewise, in stage TE, the fixed amount $k = -5$ is replaced by another ballot box with ten balls that represents the other zero-mean risk in the definitions of A_4 and B_4. See the instructions in the Appendix A.2.2 for explicit screenshots of these stages.

2.3.3 Discussion of experimental method and parameter choices

In a theoretical paper, Ebert (2010) analyzes the statistical properties of the proper risk apportionment lotteries of Eeckhoudt and Schlesinger (2006) and shows that they are mostly driven by the skewness of the zero-mean risks that have to be apportioned. As a consequence, he shows that no binary lottery can capture the essential features of the proper risk apportionment lotteries of 3rd-order and higher sufficiently well. This is also observed in the experiment of Ebert and Wiesen (2009) who show that the skewness of the zero-mean risk indeed influences subjects' decisions significantly. Thus, we need at least trinomial lotteries. The temperance lotteries of Eeckhoudt and Schlesinger with binary zero-mean risks involve up to 5 states. Comparison between two such lotteries would pose a significant challenge in comprehensiveness to subjects. Thus, we use the representation as a compound lottery introduced in Ebert and Wiesen (2009) that they also test for experimental robustness. This representation also fits well the intuition of proper risk apportionment (disaggregation of harms across states of nature, 'putting risk in its proper place') as defined by Eeckhoudt and Schlesinger (2006).

Probabilities do not vary within one decision screen and, generally, the only probabilities used are 50/50 and 80/20. Unlike the probabilities the outcomes of the "more risky" option of a lottery pair are varied in our procedure. That it is meaningful to vary outcomes rather than probabilities to change the expected value of a risky prospect has recently been shown by Bruner (2009). The probabilities of our lottery pairs are visualized using ballot boxes similar to the ones actually used to determine subjects' payoffs (see Figure 2.4). They were shown to subjects prior to the experiment while explaining the decision screens.

The multiple price list procedure is one of the most established methods to measure individual risk attitudes. As already described in Subsection 2.3.2 the variant applied in our experiment confronts a subject with an array of ordered prices (here m^{RA}, m^{PR} and m^{TE}) and asks the subject to make a decision between Option A and Option B for each price. In general, the procedure is very attractive as it easy to implement and the task involved can be easily accessed by the subjects. Moreover, truthful revelation is in subjects' best interest. However, one frequently raised concern is that the multiple price list method is prone to framing effects, as subjects are drawn to the middle row of the ordered list irrespective of their true values; see Andersen et al. (2006) for a comprehensive discussion. To account for this possible effect we devise a test by varying the cardinal scale of the multiple price list as will be explained in Subsection 2.3.4. Notice however that our qualitative results are unaffected by this framing issue.

Our approach is largely inspired by the theoretical paper of Crainich and Eeckhoudt (2008). However, the downside risk premium m^{PR} they define is added to A_3 in the "good" state only, while we define it as being added to both states of the 50/50 gamble. That is, we consider $A_h + m_h$ ($h = 2, 3, 4$). A risk premium for sure seems to be more comprehensive for experimental purposes than a risk premium 'with probability'. Further, crucial to our approach is that this simplifies the calibration of the lotteries significantly because A_h

differs from $A_h + m_h$ only by its mean while all higher central moments are unaffected by the size of the risk premium. This allows for a "clean" tradeoff between increased mean and 2nd-degree risk, downside risk or outer risk, respectively. Furthermore, this is why we will be able to reasonably compare the risk premia of various orders obtained in each stage. Note that, although the risk premia are added for sure, there is no experimental certainty effect because both Option A and Option B are always risky.

Moreover, our approach is not based on moments. It is, rather, insightful to look at the

Table 2.1: Statistical features of lottery pairs employed in the experiment

	Stage RA	Stage PR			Stage TE	
		PR1	PR2	PR3	TE1	TE2
$E[B_h] - (E[A_h] + m_h^{\cdot})$	0.00	0.00	0.00	0.00	0.00	0.00
$V(B_h) - V(A_h)$	50.00	0.00	0.00	0.00	0.00	0.00
$Skew(B_h) - Skew(A_h)$	0.00	2.16	2.16	2.16	0.00	0.00
$Kurt(B_h) - Kurt(A_h)$	0.00	0.00	-5.44	5.44	-1.92	-3.00

This table shows differences in the first four standardized central moments between the lottery pairs in each of the six tasks of the experiment. The risk premia $m_i^{\bullet}$ only influence expectation and thus do not distort higher-order risk features of the lotteries. The risk averse lottery choice has a smaller variance than the risk-loving lottery choice. The prudent lottery choice has a higher skewness than the imprudent choice but can have a smaller or higher kurtosis, depending on the skewness of zero-mean risk that has to be apportioned. The kurtosis is smaller if and only if the zero-mean risk is left-skewed. The temperate lottery choice has a smaller kurtosis.

moments of the particular lotteries employed in our experiment to supplement our intuition for the different orders of risk. Table 2.1 illustrates in what moments the lotteries to test for different risk orders differ. For more on higher-order risk preferences, nth-degree risk and moments see Ekern (1980), Roger (forthcoming) and Ebert (2010). The latter paper shows that, for prudence and temperance, the intuition provided by considering the first four moments only essentially generalizes to more general notions of skewness and kurtosis as defined by all odd and even moments, respectively. For example, prudence is shown to be a preference for high odd moments (skewness) irrespective of the even moments (kurtosis). Ebert and Wiesen (2009) observe a significant effect of kurtosis on prudent decision behavior and thus the present experiment comprises three tasks for prudence to respect that effect.

To facilitate comparisons *between* the premia measured in each task, all lottery pairs are calibrated according to their moments up to the third order. That is, for equal values of the risk premium, the six lotteries with less (more) nth-degree risk have expectation 17.5 ($17.5 + m_i^{\bullet}$). Because we choose the risk premia to be added for sure, they do not distort higher-order moments of the lotteries. Naturally, the risk averse lotteries differ by their variance. Independent of the value of m^{RA}, for the lotteries in task RA we have $0.5\,(V(B_2) + V(A_2)) = 31.25$. The prudence and temperance lotteries in our experiment are constructed such that $V(A_3) = V(B_3) = V(A_4) = V(B_4) \approx 31$ with an error of less than 0.25. Similarly the risk aversion and temperance lotteries have a skewness of approximately 0 which is the average skewness of the six prudence lotteries used in tasks PR1 to PR3. Thus we test for risk aversion, prudence and temperance not only in the same "wealth region", but also in the same "risk region" in terms of variance and skewness. Together with the consistent decision framing as proper risk apportionment in all three stages with the same type of decision screen, this should make the premia of different order we elicit reasonably comparable.

2.3.4 Robustness tests and factorial design

To account for possible disadvantages associated with our method to elicit subjects risk preferences, like order effects and framing, we employ a between subjects 2^4-factorial design, i.e., four factors (A, B, C, D) with two levels each. For all $2^4 = 16$ possible combinations of factor levels 8 subjects make their choices in the experimental stages (see Table 2.2). This explains why our experiment was outlined for $16 \cdot 8 = 128$ subjects in total in the experiment. The sequence of sessions (with their particular factor constellations) was randomized. See Montgomery (2005) for an overview of the factorial design technique.

Within our factorial design we test for *order effects* (Factors A and B) that potentially can distort results (see, e.g., Harrison et al. 2005, Andersen et al. 2006). As shown in Table 2.2, Factor A implies that stage TE is either the first or last stage a subject enters and Factor B is either "stage RA precedes stage PR" or "stage PR precedes stage RA". Thus, we consider four out of six possible stage sequences: TE-PR-RA, TE-RA-PR, PR-RA-TE, RA-PR-TE. Note that the sequences of tasks in stages PR and TE is randomized individually.

The multiple price list instrument might suggest a frame that encourages subjects to select the middle row of the m-list to switch from one lottery to the other, contrary to their unframed risk preferences; see, e.g., Andersen et al. (2006), Harrison et al. (2007). To account for this potential problem we devise a *shifted grid* (Factor C) in which the middle row implies different risk attitudes. More specifically, the levels of Factor C are either a shift in the scale of the m-list (2.00 EUR are added to each value) or no shift. We also deliberately changed the grid distances of the m-list in order to detect behavioral influences

Table 2.2: Factorial design

Ses.	Level of Factor A	Level of Factor B	Level of Factor C	Level of Factor D	Number of subjects (female, male)
1	TE last	PR-RA	no Shift	Grid 0.50	8 (4,4)
2	TE last	RA-PR	no Shift	Grid 0.25	8 (4,4)
3	TE last	RA-PR	no Shift	Grid 0.50	8 (4,4)
4	TE last	PR-RA	no Shift	Grid 0.25	8 (4,4)
5	TE first	PR-RA	Shift	Grid 0.50	8 (4,4)
6	TE last	RA-PR	Shift	Grid 0.50	8 (4,4)
7	TE first	RA-PR	Shift	Grid 0.50	7 (4,3)
8	TE last	RA-PR	Shift	Grid 0.25	8 (4,4)
9	TE first	RA-PR	no Shift	Grid 0.25	8 (4,4)
10	TE first	PR-RA	no Shift	Grid 0.50	8 (4,4)
11	TE first	PR-RA	Shift	Grid 0.25	8 (4,4)
12	TE first	RA-PR	Shift	Grid 0.25	8 (4,4)
13	TE first	PR-RA	no Shift	Grid 0.25	8 (4,4)
14	TE last	PR-RA	Shift	Grid 0.50	8 (4,4)
15	TE first	RA-PR	no Shift	Grid 0.50	8 (4,4)
16	TE last	PR-RA	Shift	Grid 0.25	8 (4,4)

This table illustrates the run-order of the sixteen sessions we conducted, the factor constellation for every session and the number of participants and their gender. In every session we collected responses of 4 men and 4 women, except for session 7, where only 3 women (including the invited substitutes) showed up.

of *grid increments* (Factor D), i.e., we used a fine and a coarser grid. Factor D considers the distance between two values on the grid being either 0.25 EUR (such as described in the previous subsections) or 0.50 EUR. Consequently, depending on the levels for Factors C and D the m-list adapts four different ranges: (i) $[-2.50, 2.25]$ and (ii) $[-0.50, 4.25]$ both with a gird of 0.25 EUR as well as (iii) $[-5.00, 4.50]$ and (iv) $[-3.00, 6.50]$ with a grid of 0.50 EUR.

Gender differences in risk preferences, i.e., risk aversion, is a well documented phenomenon in the experimental economics literature (Croson and Gneezy 2009). To consider possible gender effects in the decision tasks of our experiment the number of male and female participants is well balanced for each session.

2.3.5 Experimental procedure

The experimental sessions were conducted at Bonn$E$con$L$ab in January and February 2010. Overall 127 students from various disciplines, e.g., mathematics, economics, law, business administration, history and linguistics, participated in our 16 experimental sessions. Subjects were recruited by the online recruiting system ORSEE (Greiner 2004). As already shown in Table 2.2 the number of male and female participants was well balanced in each session. The experimental sessions lasted for about 1.5 to 2 hours. Subjects earned on average 24.00 EUR.

The procedure of the experiment was as follows: Firstly, experimenters extensively introduced the decision task and the procedure of the experiment to the subjects. Before each experimental stage started, subjects were asked to answer control questions testing their understanding of the decision task. In particular, they were familiarized with the illustration of lotteries and their outcomes as well as probabilities. Only when subjects had answered the questions correctly they were allowed to proceed to the experimental decisions. Then subjects made the decisions in the experimental stages. Fourthly, subjects answered the questionnaire comprising demographic questions. For answering the questionnaire subjects received 4.00 EUR in addition to their earnings from the experiment.

Finally, each subject's payoff was determined by a random-choice payment technique. As already mentioned, subjects made a series of 120 choices, each with substantial monetary consequences, and final payoff are based on just *one* of these choices selected at random after all have been completed. The random choice was made by drawing one card out of a set of cards numbered between 1 and 120 from a ballot box. The randomly drawn choice could either be from stage RA, PR or TE. Moreover, the lottery of the randomly chosen question determined by a subject's choice is actually played out. Corresponding to the question randomly chosen a coupon was allocated to the lottery outcome. Afterwards the experimenters gave the resulting payoff to the subjects.

2.4 Behavioral results

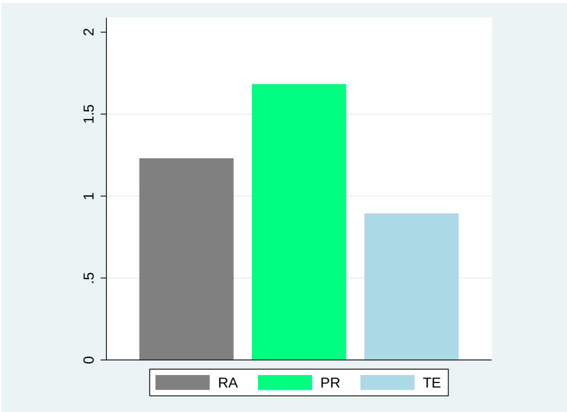
In this section, we present the results from our experimental sessions. Firstly, we report risk taking behavior on the aggregate for each risk type. Secondly, we explore the relationship between the different risk types. Thirdly, the robustness of our experimental method is checked and, finally, we analyze risk taking behavior across gender.

2.4.1 Premia for different risk types

In all questions, the vast majority of subjects chose the "less risky option" when compensation was small, and then crossed over to the "more risky option" without ever going back to the less risky option. Aggregated over six tasks, 85% of individuals switched once and 8% did not switch. For 3% we observe two switches and in about 4% of responses subjects switch back and forth more than two times. This latter fraction is slightly lower than reported in other multiple price list experiments to elicit risk preferences, e.g., Holt and Laury (2002), who report between 5 to 6% of multiple switches. We included subjects in our analysis with one switching point or no switch at all. Further, we include subjects with two switching points. We dropped subjects from our analysis who had more than two switching points for more than *one* out of the six decision screens. This was the case for eight out of 127 subjects.

In the following, an individual's *response to a task* refers to the first premium for

Figure 2.5: Average premium by risk type



This figure shows the average desired premia of 119 subjects by risk type, $\overline{\hat{m}}^{\mathrm{RA}}$, $\overline{\hat{m}}^{\overline{\mathrm{PR}}}$ and $\overline{\hat{m}}^{\overline{\mathrm{TE}}}$.

which an individual switched to the more risky lottery choice in that task. A subject's *response to a stage* is the average of the subject's responses to the tasks of that stage. Formally, let $\hat{m}_i^{\mathrm{RA}}$ denote individual i's response in stage RA (which consisted of one task only). Further, let $\hat{m}_i^{\overline{\mathrm{PR}}} := \frac{1}{3}\left(\hat{m}_i^{\mathrm{PR1}} + \hat{m}_i^{\mathrm{PR2}} + \hat{m}_i^{\mathrm{PR3}}\right)$ and $\hat{m}_i^{\overline{\mathrm{TE}}} := \frac{1}{2}\left(\hat{m}_i^{\mathrm{TE1}} + \hat{m}_i^{\mathrm{TE2}}\right)$ denote individual i's average response to the three prudence and two temperance tasks, respectively. The corresponding averages over all individuals are denoted by $\overline{\hat{m}}^{\mathrm{RA}}$, $\overline{\hat{m}}^{\overline{\mathrm{PR}}}$ and $\overline{\hat{m}}^{\overline{\mathrm{TE}}}$, respectively. These overall averages are depicted in Figure 2.5. We clearly observe that subjects desire more compensation to accept the imprudent compared to the risk loving and intemperate choice.

In particular, as shown in Table 2.3 subjects desire on average a higher compensation to accept an imprudent choice ($\overline{\hat{m}}^{\overline{\mathrm{PR}}} = 1.6817$; s.d. 1.3427) compared to the risk loving ($\overline{\hat{m}}^{\mathrm{RA}} = 1.2290$; s.d. 1.8012) and the intemperate choice ($\overline{\hat{m}}^{\overline{\mathrm{TE}}} = 0.8929$, s.d. 1.2175).[8]

[8]Analyzing the medians for m_i^{RA}, $\hat{m}_i^{\overline{\mathrm{PR}}}$ and $\hat{m}_i^{\overline{\mathrm{TE}}}$ clearly indicates a tendency towards risk averse, prudent and temperate behavior, as subjects' responses differ substantially from the risk neutral choice, i.e., the choice of a premium of 0 (or put differently, crossing over from the less risky to the more risky choice when the expected value of more risky choice is larger for the first time). The median premia for the risk loving, imprudent and intemperate choice are 1.00, 1.50 and 0.50, respectively. Note that premia for

This behavioral pattern implies that subjects attach, on average, more weight to third-order risks than to second-order and fourth-order risks. We also observe this pattern for subjects' responses per task. Table 2.3 shows that a higher premium is desired for all three tasks in stage PR compared to stage RA and the tasks in stage TE. Further, the average compensation to accept the risk loving choice is larger than the desired compensation for the two temperance items.

To test these differences for significance, we first conduct a Page-Test for ordered

Table 2.3: Descriptive statistics on subjects' desired premia

	$\hat{m}_i^{\mathrm{RA}}$	$\hat{m}_i^{\mathrm{PR1}}$	$\hat{m}_i^{\mathrm{PR2}}$	$\hat{m}_i^{\mathrm{PR3}}$	$(\hat{m}_i^{\overline{\mathrm{PR}}})$	$\hat{m}_i^{\mathrm{TE1}}$	$\hat{m}_i^{\mathrm{TE1}}$	$(\hat{m}_i^{\overline{\mathrm{TE}}})$
Mean	1.2290	1.8361	1.6940	1.5192	1.6817	0.9916	0.8098	0.8929
s.d.	1.8012	1.7837	1.6142	1.6325	1.3427	1.4287	1.4221	1.2175
Median	1.00	2.00	1.50	1.50	1.5	0.5	0.5	0.5
N	119	119	116	117	359	119	117	227

This table shows descriptive statistics on premia for each tasks and averages over tasks for stages PR and TE.

alternatives. The null hypothesis is that on average subject's responses were the same in every stage and the alternative hypothesis is that they are ordered in a *specific* way.[9] We suppose $\hat{m}_i^{\overline{\mathrm{PR}}} \geq \hat{m}_i^{\mathrm{RA}} \geq \hat{m}_i^{\overline{\mathrm{TE}}}$ to be the *specific order*. The null hypothesis of equality of responses can be rejected and, thus, it follows that *at least one* of the differences is a strict inequality ($p = 0.0004$, $L = 1480$, Page-test).[10]

Pairwise comparisons were conducted using a two-sided Wilcoxon signed rank (WSR) test and a t-test for paired samples (t). The normality assumption of the t-tests should be well satisfied given our sample size. The null hypothesis that the premia for risk aversion and prudence, $\hat{m}_i^{\mathrm{RA}}$ and $\hat{m}_i^{\overline{\mathrm{PR}}}$, have the same mean, i.e., $\overline{\hat{m}}^{\mathrm{RA}} = \overline{\hat{m}}^{\mathrm{PR}}$ is rejected ($p = 0.0057$, t and $p = 0.0101$, WSR). Likewise, the average compensation to accept the imprudent choice ($\overline{\hat{m}}^{\mathrm{PR}}$) differs significantly from the average compensation for the intemperate choice $\overline{\hat{m}}^{\mathrm{TE}}$ ($p = 0.0000$, t and WSR). The p-values for the null hypothesis that the means of the 2nd-degree risk premia $\overline{\hat{m}}^{\mathrm{RA}}$ and the outer risk premia $\overline{\hat{m}}^{\mathrm{TE}}$ are

all risks differ significantly from the risk neutral choice ($p < 0.0001$, Wilcoxon signed rank test, two-sided).

[9]To specify the null hypothesis and its alternative more explicitly, let $\theta(\cdot)$ be the population median of subjects' responses. Then the null hypothesis that the medians are the same may be written as $H_0 : \theta(\hat{m}_i^{\mathrm{RA}}) = \theta(\hat{m}_{\mathrm{PR}}^{i}) = \theta(\hat{m}_i^{\mathrm{TE}})$ and the alternative hypothesis may be written as $H_0 : \theta(\hat{m}_i^{\overline{\mathrm{TE}}}) \leq \theta(\hat{m}_i^{\mathrm{RA}}) \leq \theta(\hat{m}_i^{\overline{\mathrm{PR}}})$ where at least one of the differences is a strict inequality. That is, the medians are ordered in magnitude. Notice that, we corrected for ties.

[10]Notice that if we assume the ordering $\hat{m}_i^{\mathrm{RA}} \geq m_i^{\mathrm{PR}} \geq \hat{m}_i^{\mathrm{TE}}$ the null hypothesis cannot be rejected ($p = 0.4364$).

$p = 0.0527$ (t) and $p = 0.2024$ (WSR). As application of the stronger t-test seems justified, we conclude that there is weak evidence that 2nd-order risk premia are higher than outer risk.

Result 1. *On average, subjects desire significantly higher (third-order) downside risk premia than second-order and outer risk premia. We further find weak evidence that second-order premia are higher than outer risk premia.*

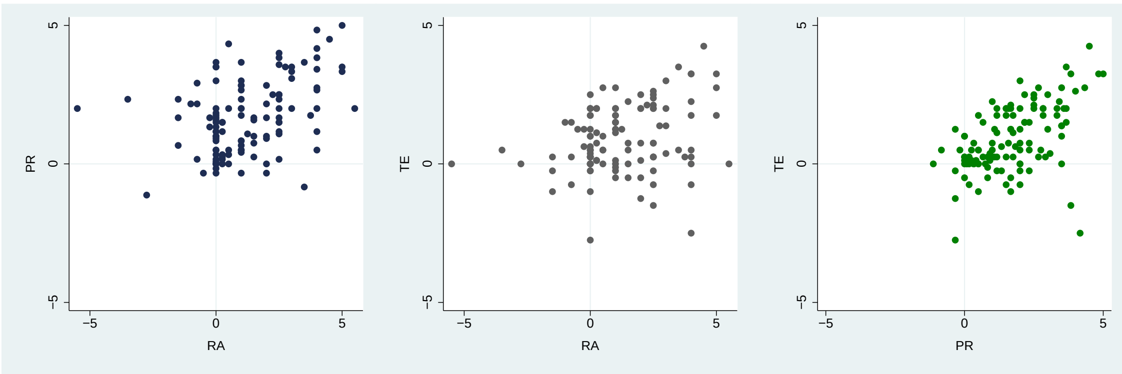
Result 1 is the main result of this paper. It shows that in a direct comparison, prudence is relatively more important to subjects than risk aversion. This is meant in the sense that they are willing to pay a larger premium to get a lottery with less downside risk compared to getting a lottery with less 2nd-degree risk. This questions the extensive focus on risk aversion in the economic literature, both theoretical and empirical (experimental), and highlights the importance of prudence. In particular, the experimental economic literature contains numerous different methods to measure risk aversion, but this paper constitutes the first approach to measure prudence. When comparing two risks (rather than comparing risk with certainty), risk aversion no more exhaustively specifies subjects' risk preferences. To do this, risk aversion must be complemented by higher-order risk preferences. As shown here, a 2nd-order risk increase which is addressed by risk aversion is not the most important one to subjects. A downside risk increase is more harmful and this will be reflected by an individual's preferences if and only if prudence is assumed.

The only significant difference for average premia *within stages PR and TE,* respectively, is observed for PR1 and PR3 ($p = 0.0269$, t and $p = 0.0561$, WSR). This confirms that subjects indeed should respond to several prudence tasks as the choice of the zero-mean risk influences decision behavior (see also Ebert and Wiesen, 2009). In particular, if we had only employed stage PR3 we would not have observed a significant difference in premia to stage RA (0.1331, t and 0.3013, WSR). The observation that the imprudence premium is smallest for a right-skewed zero-mean risk as employed in task PR3 seems reasonable as such a risk constitutes less of a harm to a prudent individual. It can further be shown that in this case prudence implies choosing higher kurtosis (as defined by *all* even moments being higher) what might make the prudent option less attractive to a temperate individual, see Ebert (2010). Further, according to that paper, a mixed risk averse decision maker should demand premia that are increasing in the skewness of the zero-mean risk. We find some significant support for this in the comparison of PR1 to PR3. Further, the premium in PR2 is higher than in PR3, but not significantly. Contradictory to mixed risk aversion is that the premium in PR1 is higher than in PR2 (but not significantly). Let us finally also note that Maxmin preferences (e.g., Gilboa 2009, chapter 17) cannot explain our behavioral result as the temperance lotteries involve the highest losses but the corresponding outer risk premia are smallest.

2.4.2 Relationship between risk aversion, prudence and temperance

Theoretically, risk aversion, prudence and temperance are complementary in describing individuals' risk attitudes. But what is the relationship empirically? In the following we explore this question by analyzing each individual's demanded premium for the three different risk types. Figure 2.6 shows three scatter plots contrasting individual premium choices for risk types in a pairwise manner. For all three comparisons the plots suggest a positive correlation. Test statistics of a Spearman rank correlation test confirm this

Figure 2.6: Pairwise comparison of premia for different risk types



The left graph plots jointly the risk premia demanded for prudence (vertical axis) and risk aversion (horizontal axis) by each of 119 individuals. The centered (right) graph plots the premia for temperance and risk aversion (temperance and prudence).

tendency. There is a significant positive relationship between the second-order risk premium $\hat{m}_i^{\mathrm{RA}}$ and the downside risk premium $\hat{m}_i^{\mathrm{PR}}$ ($r_s = 0.3896$, $p = 0.0000$). Moreover, the correlation between $\hat{m}_i^{\mathrm{RA}}$ and $\hat{m}_i^{\mathrm{TE}}$ is also positive ($r_s = 0.2681$) and significant ($p = 0.0032$). The 'strongest' positive relationship can be observed between responses in stage PR ($\hat{m}_i^{\mathrm{PR}}$) and stage TE ($\hat{m}_i^{\mathrm{TE}}$), as $r_s = 0.5805$ at a 1% significance level.[11] This supports the assumption of mixed risk aversion which is common in the economic literature.

[11]The relationships are qualitatively the same when considering premia for the tasks in stages PR and TE separately, i.e., $\hat{m}_i^{\mathrm{PR1}}, \hat{m}_i^{\mathrm{PR2}}, \hat{m}_i^{\mathrm{PR3}}$ and $\hat{m}_i^{\mathrm{TE1}}, \hat{m}_i^{\mathrm{TE2}}$.

Result 2. *Behavioral data evidence a significantly positive relationship between individuals' demanded premia for second-order risk, downside risk and outer risk. This implies that risk aversion, prudence and temperance often occur jointly (but with different intensity, see Result 1). The highest positive correlation can be observed between prudence and temperance.*

2.4.3 Robustness analysis

In this section we analyze the factorial design described in Section 2.3.4. We test robustness of our experiment towards stage order effects (Factors A and B) and manipulations of the premia grid. For Factor C the levels are either "the premia grid is shifted by 2.00 EUR" or "the premia grid is not shifted". Depending on Factor D the grid size was either 0.25 EUR or 0.50 EUR.

At first we analyze whether subjects' average premia over all six tasks, $\overline{m}_i := \frac{1}{6}(\hat{m}_i^{\text{RA}} + \hat{m}_i^{\text{PR1}} + \hat{m}_i^{\text{PR2}} + \hat{m}_i^{\text{PR3}} + \hat{m}_i^{\text{TE1}} + \hat{m}_i^{\text{TE2}})$, varies for different factor levels. Table 2.4 shows descriptive statistics by factor levels and provides p-values of a two-sided Fisher-Pitman permutation test.

The order in which stages occur does not significantly influence subjects' responses (A: $p = 0.4126$, B: $p = 0.1271$). However, shifting the scale of the premia by +2.00 EUR does significantly influence subjects responses. When there is no shift subjects' desired premia are lower than when there is a shift. This difference is significant at a 1% level (C: $p = 0.0077$). Although the average responses are slightly larger for a grid of 0.50 EUR on the premium scale than for a grid of 0.25 EUR, the change in grid size did not exert a significant influence on subjects' decisions (D: $p = 0.6104$).[12]

Are the same factors still influential for subjects' behavior when we distinguish between responses for individual the stages RA, PR and TE, i.e., $\hat{m}_i^{\text{RA}}$, $\hat{m}_i^{\overline{\text{PR}}}$ and $\hat{m}_i^{\overline{\text{TE}}}$? Different levels of Factors A and D caused no significant influence on the premia for the three different types of risk. For Factor B the intemperance premium is significantly larger when stage RA precedes PR. Subjects desire a substantial higher premium for all types of risk when there is a shift in the scale of premia (Factor C). This difference is significant for $\hat{m}_i^{\text{PR}}$ and $\hat{m}_i^{\text{TE}}$. However, for $\hat{m}_i^{\text{RA}}$ the difference is substantial but not significant ($p = 0.11263$). To sum up, as is typical for experiments employing a multiple price list format, shifts in the premia grid can potentially distort measurements such that one should control for this

[12]p-values from a parametric t-test for unpaired samples are very similar; i.e., A: $p = 0.4425$, B: $p = 0.1260$, C: $p = 0.0071$ and D: $p = 0.6081$.

Table 2.4: Analysis for different factor levels

Factor (level)		$\overline{m}_i$	m_i^{RA}	$\hat{m}_i^{\mathrm{PR}}$	$\hat{m}_i^{\mathrm{TE}}$
A (TE first)	Mean	1.4244	1.4831	1.5367	0.7288
	s.d.	1.1202	1.9208	1.3310	1.2834
A (TE last)	Mean	1.2579	1.6534	1.8243	1.0542
	s.d.	1.0834	0.9792	1.3499	1.1367
	p	0.4126	0.1311	0.2439	0.1482
B (PR-RA)	Mean	1.5004	1.1353	1.5731	0.6455
	s.d.	0.9959	1.9146	1.4049	1.3410
B (RA-PR)	Mean	1.1911	1.3276	1.7960	1.1530
	s.d.	1.1802	1.6847	1.2761	1.0206
	p	0.1271	0.5742	0.3672	0.0226
C (no Shift)	Mean	1.0794	0.9713	1.3805	0.6742
	s.d.	1.0207	1.6552	1.2604	1.2540
C (Shift)	Mean	1.6180	1.5000	1.9987	1.1228
	s.d.	1.0207	1.9200	1.3641	1.1439
	p	0.0077	0.1126	0.0122	0.0443
D (Grid 0.25)	Mean	1.3926	1.2629	1.5611	0.8966
	s.d.	1.2170	1.3990	1.1441	0.9954
D (Grid 0.50)	Mean	1.2885	1.1967	1.7965	0.8893
	s.d.	0.9712	2.1258	1.5082	1.4050
	p	0.6104	0.8538	0.3437	0.9828

This table shows descriptive statistics on premia averaged over risk types $\overline{m}_i$ and on premia per risk type for each factor level. Further it shows p-values of a two-sided Fisher-Pitman permutation test for independent samples.

effect.

The more important point with respect to the application of the factorial design is the following. The factorial design introduced a lot of variation into our measurements, but still we obtain significance for our results. That is, these results are robust towards caveats of the experimental method.

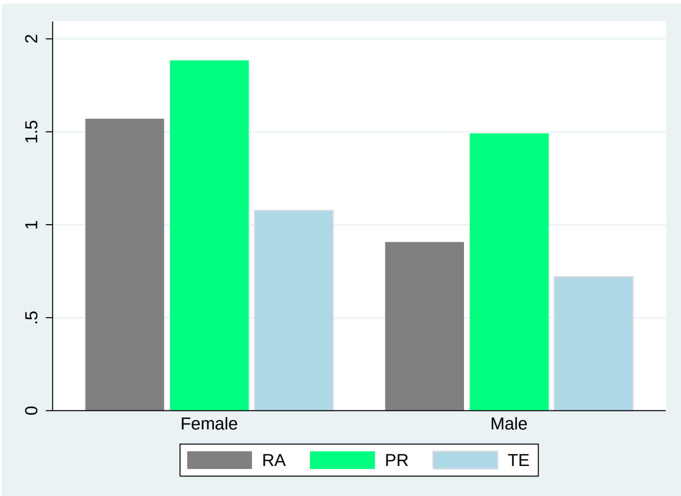
Result 3. *The order of stages does not influence subjects' average premium choice significantly. Also the grid increments do not influence subjects' choices. A shift in the premia grid influences subjects' behavior for all orders of risk significantly. It is important to note that the other results of the experiment are significant despite of being challenged by the factorial design.*

2.4.4 Is there a male-female difference?

Differences between women and men in risk attitudes are well documented in the experimental economics literature. Most evidence suggests that women perceive risks as greater, engage in less risky behavior, and choose alternatives that involve less risk. In their literature reviews Eckel and Grossman (2008 b) and Croson and Gneezy (2009) conclude that it is a robust finding from (economic) experiments that women are more risk averse than men. In this section we show that this observation also applies to the higher-order risk preferences prudence and temperance.

Figure 2.7 illustrates average demanded compensations for the different types of risk for 58 women and 61 men. The finding that a higher premium is desired for the imprudent choice than for the risk loving and intemperate choice is robust for both males and females. In line with the literature on gender differences in risk taking behavior we find that

Figure 2.7: Average compensation by gender



This figure shows the average desired premia of 119 subjects for risk types for female and male subjects, i.e., $\overline{\hat{m}}^{\text{RA-F}}$, $\overline{\hat{m}}^{\text{PR-F}}$ and $\overline{\hat{m}}^{\text{TE-F}}$ and $\overline{\hat{m}}^{\text{RA-M}}$, $\overline{\hat{m}}^{\text{PR-M}}$ and $\overline{\hat{m}}^{\text{TE-M}}$.

female (F) subjects are more risk averse than male (M) subjects. To accept the risk loving choice women demand, on average, a higher premium than men ($\overline{\hat{m}}^{\text{RA-F}} = 1.5690$; $\overline{\hat{m}}^{\text{RA-M}} = 0.9057$). This difference is significant ($p = 0.04474$, two-sided Fisher-Pitman permutation test).[13]

Moreover, our data show that women are both more prudent and temperate than men. That is, women demand a higher compensation to accept the imprudent ($\overline{\hat{m}}^{\text{PR-F}} = 1.8829$; $\overline{\hat{m}}^{\text{PR-M}} = 1.4904$) and the intemperate choices ($\overline{\hat{m}}^{\text{TE-F}} = 1.1504$; $\overline{\hat{m}}^{\text{TE-M}} = 0.7350$). Both differences are significant at a 5% level ($p = 0.0448$; $p = 0.0432$). This puts the robust finding that "men are more risk prone than women" (Croson and Gneezy, 2009, p.449) on

[13]Notice that we employ a permutation test for *paired samples* as session averages are compared by gender.

a broader basis.

Result 4. *Women not only are more risk averse than men, but also are more prudent and more temperate.*

2.5 Conclusion

In this paper we propose an experimental method to measure risk aversion, prudence and temperance at an individual level. Within the scarce empirical literature on higher-order risk preferences, this constitutes the first attempt to measure the *intensity* of risk preferences rather than only their *direction.* Further, it is the first attempt to *compare* the intensity of the preferences within subjects.

The theoretical fundament of our experimental method is the proper risk apportionment model of Eeckhoudt and Schlesinger. Within this model we define risk premia of higher-orders and show that they are related to higher-order intensity measures in the spirit of Arrow-Pratt. By definition, the premia imply a clean tradeoff between mean and 2nd-degree risk (downside risk and outer risk, respectively) for the lottery choices in the experiment. Lotteries are calibrated such that these premia are comparable not only between but also within subjects.

In the experiment we measure these risk premia using a multiple price list technique. The lotteries employed in the experiment are presented as compound rather than trinomial (quadri- or pentanomial) and match the intuition of proper risk apportionment. The only probabilities used are 50/50 and 80/20. These probabilities are visualized using ballot boxes similar to the ones actually used to determine subjects' payoffs. This experimental design is tested for robustness to typical manipulations using a between subject factorial design.

Our main result is that the downside risk premium demanded is significantly higher than the second-order risk premium. This highlights the importance of prudence and questions the extensive focus in the economics literature on risk aversion. In particular, the literature contains numerous different experimental methods to measure risk aversion, but this paper constitutes the first approach to measure prudence.

Behavioral data implies that the outer risk premium is smallest. It is smaller than the second degree premium with weak significance and smaller than the downside risk premium with strong significance. We also observe that the stylized fact that women are more risk averse than men extends to risks of higher orders. That means, women are significantly more prudent and more temperate than men.

Further research on the measurement of higher-order risk preferences seems to be desirable in order to close the significant gap to the experimental literature on risk aversion. Given the observation that the intensity of downside risk aversion is higher than for risk aversion, this seems to be even more justified. Moreover, our method could find application in further experiments to test the predictions of numerous theoretical papers on higher-order risk preferences.

Chapter 3

How payment incentives affect physicians' provision behavior

3.1 Introduction

A central concern in health economics is to understand the influence of institutions on the behavior of actors on health care markets. In practice, effects from changing institutions like the payment system during a health care reform are *ex ante* not necessarily known to policy makers and may influence behavior in an undesired way. Main addressees of reforms are health care providers (physicians) whose behavior is likely to be influenced by the payment system. Theoretical health-economic literature has highlighted the different incentives of commonly used payment systems like fee-for-service (FFS) or capitation (CAP). Under FFS physicians are paid for each medical procedure or service dispensed to a patient whereas under CAP, physicians receive a fixed payment for each patient irrespective of the quantity of medical services provided. FFS inherits an incentive to 'overserve' patients whereas CAP may lead to underprovision of medical services (Ellis and McGuire 1986, Newhouse 2002).

Field studies show that different payment systems do affect physicians' behavior. Yet, the results are often not comparable because of country-specific institutional differences.[1] In some studies, more than one component of the payment system is varied simultaneously making causal inferences difficult or even impossible. According to Gosden et al. (2001) the results are too contradictory to draw a definite conclusion on the direction of an effect.

[1]See for example the studies by Stearns et al. (1992) and Davidson et al. (1992) in the US, Krasnik et al. (1990) in Denmark, Iversen and Lurås (2000) and Grytten and Sørensen (2001) in Norway, Hutchinson et al. (1996), Devlina and Sarma (2008) and Dumont et al. (2008) in Canada.

Another empirical method is called for that complements field studies and overcomes (some of) the problems mentioned above. Fuchs (2000) in his article on the future of health economics argues that incorporating methods of experimental economics into health economic research may lead to great benefits for the latter. In a similar vein, Frank (2007) argues in favor of applying behavioral economics methods in health economics.

Our study contributes to the research agendas suggested by Fuchs and Frank. We use a controlled laboratory experiment to improve the understanding of the institutional parameter 'payment system' by implementing the specific features of FFS and CAP. The main focus of our study is on *how* the two payment systems influence a physician's provision of medical services, and we abstract from factors other than the payment system. Our study is one of the very first ones tackling a health economic topic by methods of experimental economics.[2]

In our experiment, medical students in the role of physicians decide on the quantity of medical services under the two payment systems. Patients gain a benefit from these services—the patient benefit measured in monetary terms. Only abstract patients 'participate' in our experiment. To provide the physicians in the experiment with an incentive for favorable behavior towards the patients, however, the money corresponding to the benefits of all abstract patients is transferred to a charity caring for real patients.

Our main finding is that physicians are influenced by the payment system. In line with theoretical considerations, patients are overserved under FFS and underserved under CAP. Financial incentives are not the only motivation for physicians' quantity decisions. The patient benefit is of rather considerable importance as well. Patients are affected differently by the two payment systems. Those in need of a low level of medical services are better off under CAP, whereas patients in need of a high level of medical services gain a higher health benefit when physicians are paid by FFS.

This chapter is organized as follows. Section 3.2 sketches the theoretical and empirical literature on physician payment and incentives most relevant to our research topic. Section 3.3 states our research questions. Experimental design and procedure are described in Section 3.4. Section 3.5 provides a statistical analysis of subjects' behavior within and across payment systems. Section 3.6 concludes.

[2]The only other studies we are aware of are Fan et al. (1998), Ahlert et al. (2008) and our related own study Hennig-Schmidt and Wiesen (2010); for the latter see Chapter 4.

3.2 Related literature

In the health economics literature, several authors have highlighted the different incentives in commonly used payment systems like fee-for-service (FFS) or capitation (CAP). In their seminal article, Ellis and McGuire (1986) let the physician (she) decide on the quantity of medical services as an agent of the patient (he) and the hospital. The physician's utility derives from two elements—the hospital's profit and the patient's benefit. According to Newhouse (2002), Ellis and McGuire's model is also applicable to a primary care setting rather resembling the setup we are interested in. This implies that the physician is assumed to be concerned about her own profit π and the patient benefit B, both depending on the quantity of medical services q. A major argument for including B into the physician's utility function is the professional code of medical ethics the physician is obliged to (Hippocratic Oath).[3] Ellis and McGuire find that FFS provides an incentive to overserve patients whereas CAP may lead to underprovision of medical services. Moreover, capitation payments can cause underprovision of *necessary* services (Blomqvist, 1991) and may lead to cream-skimming of patients (Newhouse, 1996 and Barros, 2003).

A rich empirical literature has studied various aspects of the relationship between the method of physician remuneration and physician behavior. Some empirical evidence suggests that physicians do respond to financial incentives. Krasnik et al. (1990) in a before-and-after study, analyze behavior of general practitioners in Denmark when the system is varied from a (pure) lump-sum payment to CAP supplemented by a FFS component. They find diagnostic and curative services to increase and the number of referrals to secondary care and hospitals to decrease. Concerning referral rates, Iversen and Lurås (2000) arrive at a similar result. They analyze referrals from primary to secondary care revealed by Norwegian general practitioners when the payment system was changed from a practice allowance component[4] complemented by a FFS-payment to a CAP-system with a lower FFS-component. The authors find referrals to be larger under CAP (with FFS-component) compared to FFS (with practice-allowance component). The increase in referrals may, however, not only be attributable to CAP but rather to the lower FFS-component.

In a randomized controlled study, Davidson et al. (1992) investigate behavior of office-based primary care physicians under a FFS system with high and low fees and a CAP

[3] See also Arrow (1963) who emphasized the importance of professional ethics; treatment should be determined by objective needs and not be limited by financial considerations.

[4] A practice allowance is a fixed sum of money Norwegian physicians are paid when contracting with the regional government.

system. Patients were children enrolled in the US-*Medicaid* program. Here, the frequency of primary care visits in the high FFS group was higher than in the CAP group. Apparently, CAP physicians constrain the quantity of medical services in order to reduce their costs. The fundholding regulation[5] in CAP may explain the lower referrals to secondary care as the responsibility for children's medical cost seems to outweigh the incentive to minimize cost in CAP.

In a more recent study, Dumont et al. (2008) analyze data on physician services from the Canadian province Quebec before and after a variation from FFS to a mixed system with a base wage, independent of services provided, and a reduced FFS payment. Their results suggest that physicians did react to payment incentives by reducing the volume of (billable) services under the mixed remuneration system. Moreover, these physicians increased the time spent per service and per non-clinical service. The latter is important to insure the quality of health care but is not remunerated under FFS. The results of Dumont et al. suggest a quantity-quality substitution in health care provision.

One of the most important if not the only *controlled field experiment* in health economics is the RAND health insurance study (Newhouse and the Insurance Experiment Group 1993). The main goal of this experiment was to investigate the influence of the insurance system (patients' co-payment vs. free care) on patients' health care service use and their health status. It was found that all types of services analyzed in the study fell with cost sharing but the reduced service use had nearly no adverse effect on health for the average person. Health among the sick poor was adversely affected, however. A smaller part of the study was devoted to analyzing the influence of the payment system. To this end, the authors compared the use of services under fee-for-service remuneration with that in a capitated staff model HMO (Health Maintenance Organization).[6] Cost savings were found to be noticeable, in particular due to lower hospital admission and lower estimated expenditure.

Not all studies support the strong link between physicians' payment systems and their behavior, however. For example, Hutchinson et al. (1996) do not find differences when comparing hospital utilization rates in Ontario (Canada) under FFS and CAP. For data from Norwegian physicians, Grytten and Sørensen (2001) find that after controlling for characteristics of patients and general practitioners the effects of physicians' payment sys-

[5]Such a fundholding system has the following characteristics: i) the financial resources for each patient are held in a fund and ii) the general practitioner is usually the decision-maker for allocating the funds.

[6]In a staff model HMO physicians typically work on a salary basis.

tems is rather small.

What can be concluded from the empirical literature? Based on their meta-study, Gosden et al. (2001) acknowledge some empirical evidence that the payment system affects physician behavior. They stress, however, that field studies face various difficulties like multiple and unobservable influences on physicians' decisions, context and country-specific payment system variations that make the generalization of results difficult. In addition, several field studies suffer from methodological problems when for instance more than one component of the payment system is varied simultaneously. We will return to these issues in the next section.

3.3 Research questions

Our main research goal is to improve the understanding on how the institutional parameter 'payment system' influences physicians' behavior. To this end, we make use of experimental economics methods by running a controlled laboratory experiment.

Experimental economics is a valid research technique that can successfully complement field and survey studies. It has a variety of advantages compared to the latter methods (see Davis and Holt 1993, Falk and Fehr 2003). Experimental data is created under controlled conditions. It is gathered in experimental sessions in which human subjects supplied with *monetary incentives*[7] make *real* decisions in economically relevant decision situations. Experimental conditions and variables of interest can be varied in a controlled manner. Exogenous *ceteris paribus* variations (e.g., of the payment system) can be easily implemented. Therefore, changes in behavior can be attributed to these modifications. Different experimenters can repeat the same experiment under comparable conditions to test for the robustness of the results.

Contrary to laboratory data, field data are collected from a natural environment where many factors influence the variable(s) of interest in a way the researcher usually cannot control.[8] These are for instance institutional parameters, physicians' characteristics, uncertainty about the impact of medical services provided as well as patient characteristics like health status or type of insurance. Constant patient populations during a transition of payment systems is important for the validity of results but can most often not be

[7]Participants are paid because they are likely to behave differently when monetary consequences are involved as compared to hypothetical choices (see Camerer and Hogarth 1999 and Hertwig and Ortmann 2001).

[8]See, however, the RAND health insurance experiment (Newhouse and the Insurance Experiment Group 1993)

guaranteed. Also, the methodological deficiencies mentioned in the section above should not be neglected (see Gosden et al. 2001). This being said, laboratory experimentation apparently is a suitable research method to successfully complement theoretical and other methods of empirical investigation.

Despite the advantages of experimental economics, objections like non-representative student subject pools, low incentives, a small number of observations and the simple environment should be taken seriously. Yet, careful experimentation can avoid many of these problems (see Falk and Fehr 2003).

We are aware that our experiment is very simplistic; in reality, a physician's decision situation is much more complex. Yet, as the goal of the present study is to highlight fundamental consequences of the payment system for physicians' behavior we think simplicity to be an advantage rather than a deficiency. The focus of our study is on how the pure payment systems FFS and CAP influence an experimental physician's provision of medical services. We incorporate the two major determinants that according to the theoretical literature influence a physician's behavior, the own profit and the patient's benefit. We also include patients with different health status, so-called patient types, to account for heterogeneity in the patient population.

Our first research question is concerned with behavior in FFS. Given our experimental parameters, do experimental physicians tend to behave according to what theory predicts in that they choose a quantity of medical services q^{FFS} larger than the patient's optimal quantity q^* if the profit-maximal quantity $\hat{q}$ exceeds q^*? Taking q^* as the benchmark for the right (best) medical treatment, we conjecture patients to be overserved under FFS.

Second, we are interested in behavior under CAP. According to predictions from theoretical models we expect patients to be underserved in that physicians choose q^{CAP} lower than q^*.

Third, we are concerned with research questions related to the consequences of *both* payment systems. How does provision behavior under CAP compare to behavior under FFS? Based on our previous conjectures, we expect experimental physicians in FFS to choose more medical services than in CAP. Moreover, does the mode of payment have an impact on whether and how experimental physicians besides their own profit take the patient benefit into account? Given the professional code of medical ethics physicians are obliged to, we expect them *not* to behave in a completely self-interested manner.

We also analyze previous questions with regard to patient types. Does the payment system affect patients with different health status differently as to physicians' treatment? If so, are there differences between FFS and CAP? We expect this to be the case. The RAND health insurance experiment (Newhouse and the Insurance Experiment Group 1993), for instance, showed certain albeit small adverse health consequences concentrated among sick people from the lowest income group.

The last research question concerns the tradeoff between own profit and patient benefit the experimental physicians are faced with. In our experiment, several Pareto-efficient quantity decisions exist for each patient. Here, physicians can neither make the patient better off without foregoing own profit nor make themselves better off without inducing a benefit loss to the patient. We are specifically interested in the following questions: Does behavior with regard to Pareto efficiency and tradeoffs vary in the two payment systems? Can a classification of behavior help us to get deeper insights into decision-making like it has helped to explain behavior in other experimental settings (e.g., Selten and Ockenfels 1998 and Fischbacher et al. 2001)?

3.4 Experimental design and procedure

3.4.1 Design and parameters

We analyze physicians' provision behavior under the two payment systems FFS and CAP. No other experimental parameter is varied. The experimental design allows for a controlled *ceteris paribus* variation and a between-subject comparison.

Each subject taking part in our experiment is allocated to a physician's role deciding on the quantity of medical services to be provided for given patients. Participants are medical students expected to become physicians in the future. We deliberately choose medical students as they most likely will identify with the decision task in our experiment. And we used a context-specific framing (see the instructions in Appendix A.3.2). Both features are important as we are interested in how subjects decide in a medical context, and identification as well as framing seems to matter for behavior.

The experiment comprises two treatments. In FFS, physicians receive a fee for each unit of medical service provided. In CAP, they are paid a lump-sum payment (capitation) per patient independent of the number of medical services they dispense. All monetary amounts are measured in Taler, our experimental currency, the exchange rate being 1 Taler = 0.05 EUR (about \$0.07).

The physicians' task is to treat patients by providing them with medical services. Patients gain a benefit from these services. The patient benefit is measured in monetary terms. Three types of patients exist. These types differ in the 'benefit functions' that relate the benefit a patient receives to the number of services a physician provides. In particular, patients differ in the number of medical services rendering the optimal treatment, i.e., maximum benefit. Patients in our experiment are abstract in that no real persons participate. Yet, to provide experimental physicians with an incentive for favorable behavior towards the patients, the money corresponding to the benefits of all abstract patients is transferred to a charity caring for real patients.

Patients are further characterized by illnesses. In FFS, it has an impact on physicians' remuneration, however, as the 'remuneration function' that relates a physician's remuneration to the number of services a physician provides is determined by the respective illness. In particular, maximum remunerations differ across the five existing illnesses. The same holds for maximum profits because the costs a physician has to bear are kept constant for all decisions and across treatments. Recall that in CAP, physicians are paid a lump-sum per patient. Therefore, neither illnesses nor the number of medical services they dispense have an impact on their remuneration.

In the remainder of this subsection we describe the experimental design in more detail. Physicians decide on the quantity $q \in \{0, 1, \ldots, 10\}$ of medical services to be provided to their patients.[9] They decide for five abstract illnesses A, B, C, D, E[10] of three patient types 1, 2, 3. Patient types differ in their benefit from medical services $(B_1(q), B_2(q), B_3(q))$. Each combination of patient type and illness represents a specific patient 1A, 1B, 1C, …, 3D, 3E (Table 3.1). By each decision ($j = 1, \ldots, 15$), physicians simultaneously determine their own profit and the benefit of a given patient. The patient is assumed to be passive and fully insured accepting each medical service chosen by a physician.

Table 3.1: Order of decisions

Decision (j)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Patient type	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3
Illness	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E
Patient	1A	1B	1C	1D	1E	2A	2B	2C	2D	2E	3A	3B	3C	3D	3E

Physicians' remuneration. In FFS, physicians receive a fee for each unit of medical

[9] The range of services physicians can choose from may be interpreted as those eligible for a patient contracting with a certain health plan.

[10] We did not specify real illnesses because this turned out not to be feasible in the experimental setup.

service provided. Fees differ across services and illnesses. As points of reference for our experimental fees we used tariffs for ophthalmologist services (like the treatment of glaucoma or cataract) taken from the German scale of charges and fees for physician services (EBM).[11] Remuneration $R(q)$ increases in the quantity of medical services chosen (see Table 3.2).

In CAP, physicians are paid a lump-sum payment R per patient independent of their

Table 3.2: Physicians' remuneration $R(q)$

		Quantity (q)										
	Illness	0	1	2	3	4	5	6	7	8	9	10
FFS	A	0.00	1.70	3.40	5.10	5.80	10.50	11.00	12.10	13.50	14.90	16.60
	B	0.00	1.00	2.40	3.50	8.00	8.40	9.40	16.00	18.00	20.00	22.50
	C	0.00	1.80	3.60	5.40	7.20	9.00	10.80	12.60	14.40	16.20	18.30
	D	0.00	2.00	4.00	6.00	8.00	8.20	15.00	16.90	18.90	21.30	23.60
	E	0.00	1.00	2.00	6.00	6.70	7.60	11.00	12.30	18.00	20.50	23.00
CAP		12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00

This table shows physicians remuneration in both treatments FFS and CAP. The payment varies for different illnesses in FFS. In CAP the lump-sum payment is 12 Taler. Notice that due to a display error on subjects' screens, physicians' remuneration for illness A at $q_j = 4$ was specified at 8.40 instead of 5.80. Physician's profits were displayed correctly, however. See the paragraph on physician's profit below.

quantity decision. To make treatments comparable, R was specified at 12 Taler in CAP which is slightly above the average maximum profit per patient a physician could achieve in FFS.

Patient benefit. Patients gain a benefit from medical services, the patient benefit $B(q)$ measured in monetary terms. Patient benefits vary across patient types. This reflects the heterogeneity of the patient population treated by a physician in the real world, e.g., with regard to a patient's health status or different severities of illness. Table 3.3 shows patient benefits $B(q)$ given the quantity of medical services provided. A common characteristic of $B(q)$ is a global optimum $q^* \in [0, 10]$. The patient optimal quantity (q^*) yields the highest benefit $B(q_j^*)$ from medical services to the patient. The patient's optimal quantity is $q_j^* = 5$ for patient type 1 ($j = 1, \ldots, 5$), $q_j^* = 3$ for patient type 2 ($j = 6, \ldots, 10$) and $q_j^* = 7$ for patient type 3 ($j = 11, \ldots, 15$). After having reached the optimum, $B(q)$ declines because providing more medical services than q^* contributes negatively to a patient's benefit at the margin. Taking q^* as the benchmark for the right (best) medical treatment, we can identify overprovision and underprovision, respectively.

[11] The German EBM lists medical services and the respective fees.

It is crucial that the experimental physicians have an incentive to take the patient

Table 3.3: Patient benefit $B(q)$

	Quantity (q)										
Patient type	0	1	2	3	4	5	6	7	8	9	10
1	0.00	0.75	1.50	2.00	7.00	10.00	9.50	9.00	8.50	8.00	7.50
2	0.00	1.00	1.50	10.00	9.50	9.00	8.50	8.00	7.50	7.00	6.50
3	0.00	0.75	2.20	4.05	6.00	7.75	9.00	9.45	8.80	6.75	3.00

This table shows the benefit patients of type 1, 2 and 3 receive depending on quantity of medical services. The patient optimal quantity q_j^* yielding the maximum benefit $B(q_j^*)$ is 5 for patients of type 1, 3 for patients of type 2 and 7 for patients of type 3.

benefit into account. Therefore, the money corresponding to the benefits of all abstract patients aggregated over all decisions of all physicians was transferred to the *Christoffel Blindenmission*—a charity caring for real patients. To verify that the money was actually transferred we applied a procedure similar to the one used in Eckel and Grossman (1996). In each session, a monitor randomly selected from the participants verify, by a signed statement, that a check for the total patient benefit is written and sealed in an envelope addressed to the charity.

Physicians' profit. Further parameters relevant for physicians' decisions are costs and profit. Like real doctors, the experimental physicians have to bear costs depending on the quantity of medical services they choose. We use a convex cost function as assumed in several theoretical models (e.g., Ma 1994, 2007 and Choné and Ma 2010). $c(q_j) = 0.1q_j^2$ $\forall q \in [0, 10], j = 1, 2, \ldots, 15$ is applied in both treatments.

Profit (remuneration minus costs) varies across illnesses in FFS because fees differ for

Table 3.4: Physicians' profit $\pi(q)$

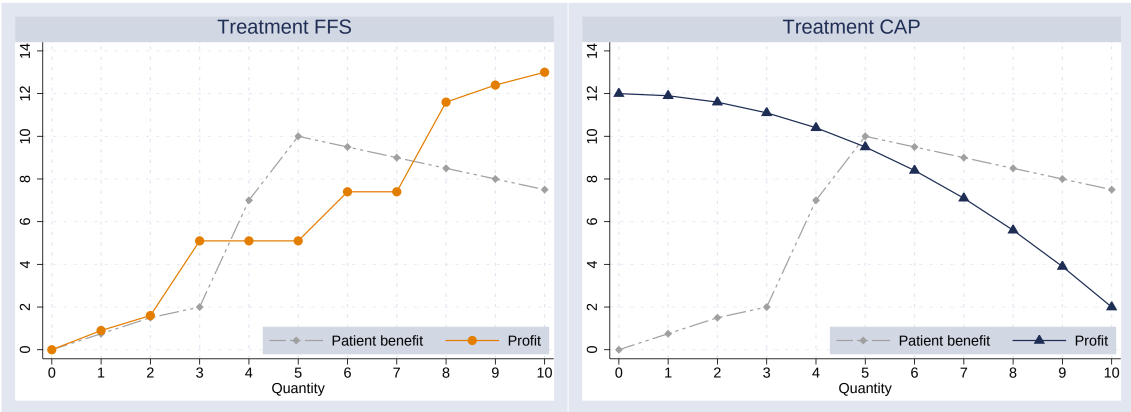
		Quantity (q)										
	Illness	0	1	2	3	4	5	6	7	8	9	10
FFS	A	0.00	1.60	3.00	4.20	4.20	8.00‡	7.40	7.20	7.10	6.80	6.60
FFS	B	0.00	0.90	2.00	2.60	6.40	5.90	5.80	11.10	11.60	11.90	12.50‡
FFS	C	0.00	1.70	3.20	4.50	5.60	6.50	7.20	7.70	8.00	8.10	8.30‡
FFS	D	0.00	1.90	3.60	5.10	6.40	5.50	11.40	12.00	12.50	13.20	13.60‡
FFS	E	0.00	0.90	1.60	5.10	5.10	5.10	7.40	7.40	11.60	12.40	13.00‡
CAP		12.00‡	11.90	11.60	11.10	10.40	9.50	8.40	7.10	5.60	3.90	2.00

‡ Physicians' maximum profit $\pi(\hat{q}_j)$ according to the profit-maximizing quantity of medical services $\hat{q}_j$.

illnesses, and costs are the same for all patients. In CAP, profit does not vary with illnesses and patient types (see Table 3.4).

For all patients in FFS, except for patient 1A ($j = 1$), the physician encounters a tradeoff between patient optimum and own profit maximization in that q_j^* differs from the profit maximizing quantity ($\hat{q}_j$). At $j = 1$ (patient 1A), $\hat{q}_j = q_j^* = 5$. For patient 3A ($j = 11$), $5 = \hat{q}_j < q_j^* = 7$. Except for illness A ($j = 1, 6, 11$) where $\hat{q}_j = 5$, the maximal profit is achieved at $q_j = 10$ (see left panel of Figure 3.1 for $j = 5$). In CAP, $\hat{q}_j = 0$ for

Figure 3.1: Patient benefit and physician's profit for patient 1E



This figure illustrates the patient's benefit and the physician's profit depending on the quantity of medical services for patient 1E (decision $j = 5$) in both payment systems FFS and CAP.

each decision $j = 1, \ldots, 15$. A higher patient benefit can only be achieved by a physician's deviating from her own maximal profit (see right panel of Figure 3.1 for $j = 5$).

3.4.2 Procedure

The computerized experiment was conducted in *B*onn*E*con*L*ab, the Laboratory for Experimental Economics at the University of Bonn. 42 medical students participated, 20 in FFS (one session) and 22 in CAP (two sessions). We thus base our analysis on 42 independent observations. Subjects were recruited by the online recruiting system ORSEE (Greiner 2004) promising a monetary reward for participation in a decision-making task. The experiment was programmed using the software z-Tree (Fischbacher 2007).

Upon arrival, participants were randomly allocated to cubicles where they took their decisions in complete anonymity. Then, subjects were provided with the instructions that were read out aloud by the experimenter. Subjects was given plenty of time for clarifying questions which were asked and answered in private. To check for subjects' understanding of the experiment we asked them to answer three test questions structured like the actual experiment but with different parameter values. The experiment was not started unless all participants had answered the test questions correctly.

The physicians in the experiment then made their 15 quantity decisions the sequence of which was predetermined, as shown in see Table 3.1, and kept across treatments. Finally, the monitor was assigned randomly. After the experiment, subjects were paid in private according to their choices. At last, the monitor verified that a check on the benefits of all patients was written and sealed in an envelope addressed to the *Christoffel Blindenmission*. The monitor and experimenter then walked together to the nearest mailbox and deposited the envelope. The monitor was paid an additional 4 EUR.

Sessions lasted for about 40 minutes. The exchange rate per Taler was 0.05 EUR. On average subjects earned 6.88 EUR in FFS and 7.42 EUR in CAP.[12] In total, 273.68 EUR were transferred to the *Christoffel Blindenmission*, 6.62 EUR per participant in FFS and 6.42 EUR in CAP. The money supported surgical treatments of cataract patients in a hospital in Masvingo (Zimbabwe) staffed by ophthalmologists of the *Christoffel Blindenmission*. Average costs for such an operation amounted to 30 EUR. Thus, the money from our experiment allowed to treat nine patients. Note that subjects were not informed about the money being assigned to a developing country (see the instructions in Appendix A.3.2).

3.5 Results

In this section we investigate physicians' behavior, both from the physician's and from the patient's perspective for FFS as well as for CAP. Moreover, we analyze the influence of physicians' profits and the patient benefit, and we study the impact of the payment system on patients' health status. We compare behavior across treatments and, finally, we analyze whether physicians' behavior is Pareto-efficient.

3.5.1 Physicians' behavior in FFS

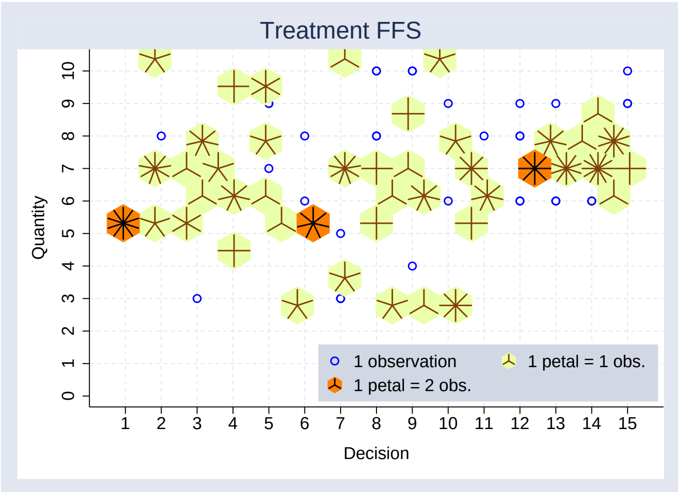
Our first research question is related to behavior under FFS. Remember that $\hat{q}_j = q_j^*$ for $j = 1$ (patient 1A), and $\hat{q}_j < q_j^*$ for $j = 11$ (patient 3A). Figure 3.2 shows absolute frequencies of of physicians' quantity choices for all patients. On average, 6.60 medical services are provided (median 7.00, s.d. 1.85). To study how patients are treated we analyze the quantity of medical services provided for each patient averaged over all physicians $\overline{q}_j = \frac{1}{20}\sum_{i=1}^{20} q_{ij}/20$.

Result 1. *In FFS, patients are overserved.*

SUPPORT: $\overline{q}_j > q_j^*$ for the 13 patients where $\hat{q}_j > q_j^*$. Patient 1A ($j = 1$) is treated

[12] Average payoffs correspond to the hourly wage of a student helper at the University of Bonn (8.32 EUR). A lunch at the student cafeteria is around 2.50 EUR.

Figure 3.2: Frequencies of quantity choices per patient in FFS



This sun flower plot illustrates the absolute frequency of physicians' quantity choices per decision (patient) in treatment FFS. The darker 'flowers' indicate two observations per petal whereas lighter flowers imply one observation per petal. Sunflowers for each decision are based on 20 observations.

optimally by all physicians i, whereas patient 3A ($j = 11$) is underserved. Testing over all patients, $\overline{q}_j$ is highly significantly larger than q_j^* ($p = 0.002$, Wilcoxon signed ranks test, two-sided). Individual physicians largely deviate from choosing the patient optimal quantities. The mean deviation from q_j^*, $\overline{\mu}_i = \sum_{j=1}^{15}(q_{ij} - q_j^*)/15$, is positive for 17 of the 20 physicians, and zero for the remainder (see Table A.3.1 in the Appendix). Thus, significantly more physicians provide medical services that are larger than q_j^* ($p = 0.003$, binomial test, two-sided). Next we investigate the impact of patient types on physicians' provision behavior.

Result 2. *Overprovision in FFS depends on patient types.*

SUPPORT: Support is provided by test statistics of an order test (see Selten 1967 and Kuon 1994) comparing the given order of average services per patient type with the perfect order (2, 1, 3) that accounts for q^* of each patient type.[13] There are six different possibilities to

[13]The logic behind the order test is the following. When a physician's quantity choice is influenced by the optimal quantity per patient types, patients in need of a large (low) quantity of medical services should on average receive a large (low) amount of medical treatment. If a physician behaves accordingly the ranks assigned to the mean quantities provided per patient type should follow a "perfect order", namely 2, 1, 3. A measure for the difference between the actual order and the perfect order is the number of inversions, i.e., the number of pairwise changes necessary to transform the given order into the perfect order. We calculate the average quantity per patient type for each of those 16 physicians whose observed order comprises three different values and rank them according to their magnitude (see Table A.3.2). For each physician, we then calculate the number of inversions necessary to achieve the perfect order of ranks.

assign three ranks. The null hypothesis of the order test is that for each subject the order of observed values is arbitrary implying the mean inversion (standard deviation) being $\mu = 1.5$ ($\sigma = 0.9574$). As we observe 0.563 average inversions only, the null hypothesis can be rejected at the 1% level.

A more in-depth analysis shows all patients of type 1 and 2 to be overserved (except for patient 1A) in that the number of physicians choosing $q_{ij} > q_j^*$ is larger than the number of physicians choosing $q_{ij} \leq q_j^*$. This is significant for four patients of type 1 and type 2 each ($p \leq 0.041$ binomial test, two-sided; see line I/FFS of Table A.3.3 in Appendix A.3.1). Patients of type 3 are treated in a less consistent way. Patient 3A (3E) is underprovided (overprovided) and the remaining patients are treated optimally by at least half of the physicians.

A physician's quantity decision determines her own profit. According to our research questions we are interested in whether profit maximizing is a main objective in general. As only 12% of the overall choices coincide with $\hat{q}_j$ this is rather not the case. Choosing $\hat{q}_j$ for all j would have yielded an average payoff of 11.08 Taler. Physicians' actual quantity decisions resulted in an average overall profit of 9.17 Taler (median 8.00 Taler, s.d. 2.69 Taler), i.e., 17% lower than $\overline{\pi}(\hat{q}_j)$. Average profits for each physician i vary between 6.53 and 10.93 Taler (see Table A.3.5).[14] Testing over all patients, $\overline{\pi}(q_j)$ is highly significantly lower than $\overline{\pi}(\hat{q}_j)$ ($p = 0.001$, Wilcoxon signed ranks test, two-sided).

We are also interested in whether profits are affected by patient types. To this end, we study the deviation of each individual physician's profit from her profit maximum, i.e., $\hat{\pi}_j - \pi_{ij}$, for patient types separately. For the sake of comparability between FFS and CAP data, we compute for each patient the *relative* deviation $\Delta\pi_{ij} = (\hat{\pi}_j - \pi_{ij})/\hat{\pi}_j$. Table A.3.4 shows $\Delta\pi_{ij}$ averaged over all physicians. Highest deviations of up to 29% are found for patients 2B and 2E, whereas lowest deviations of less than 10% occur for patients 3A and 3C. There is no deviation for patient 1A because here all physicians choose their profit maximum that coincides with the patient benefit optimum. Average profit deviation is 14.66% for patients of type 121.92% for those of type 2 and 11.98% for patients of type 3.

A physician's decision also determines the patient benefit. In FFS and CAP the benefit maximum for patients of type 3 ($B_3(q_j^*)$) is 9.45 Taler. $B_1(q_j^*) = B_2(q_j^*) = 10$ Taler (see Table 3.3). If physicians always chose the patient optimal quantity, patients would have received an average benefit $\overline{B}(q_j^*)$ of 9.82 Taler. Actual average patient benefit is 8.83 Taler

[14]For average profit per patient see Table A.3.4.

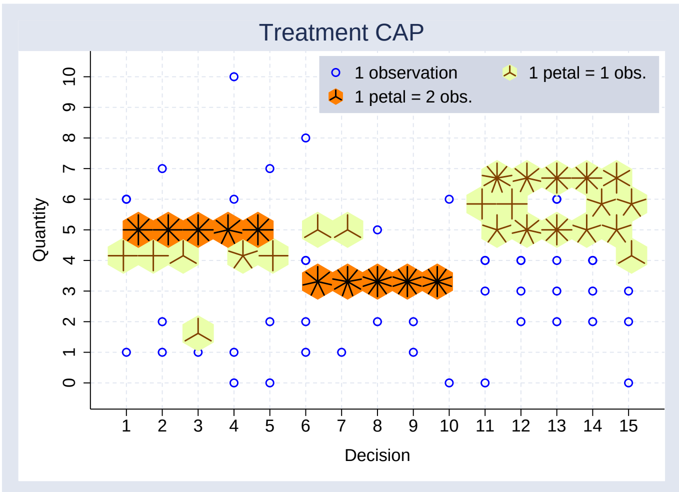
(median 9.00 Taler, s.d. 1.10 Taler), i.e., 10% lower than $\overline{B}(q_j^*)$. Further, average patient benefits determined by physician i vary between 7.52 and 9.82 Taler (see Table A.3.5).

In short, under FFS patients are overserved in that subjects on average choose quantities of medical services larger than the patient's optimal quantity. Provision is dependent on patient types as is the deviation of profits from the profit maximum. The levels of overprovision and of profit deviations tend to decrease with increasing needs of services. Physicians do not go for the maximal profit. This behavior resulted in patients receiving a substantial benefit, only 10% on average less than the maximal amount.

3.5.2 Physicians' behavior in CAP

Our second research question deals with behavior under CAP. Recall that $0 = \hat{q}_j < q_j^*$ for all patients (decisions j). Figure 3.3 shows absolute frequencies of physicians' quantity choices for all patients. On average, physicians chose 4.40 medical services (median 5.00, s.d. 1.64).

Figure 3.3: Frequencies of quantity choices per patient in CAP



This sun flower plot illustrates the absolute frequency of physicians' quantity choices per decision (patient) in treatment CAP. The darker 'flowers' indicate two observations per petal whereas lighter flowers imply one observation per petal. Sunflowers for each decision are based on 22 observations.

Result 3. *In CAP, patients are underserved.*

SUPPORT: $\overline{q}_j \leq q_j^*$ for 11 patients. Three patients (2A, 2B, 2C) are slightly overserved on average. Only patient 2E receives an optimal treatment on average. Testing over all patients, $\overline{q}_j$ is significantly smaller than q_j^* (p=0.0105, Wilcoxon signed ranks test, two-sided). Individual physicians largely deviate from the patient optimal quantity; but in contrast to FFS, they underserve in CAP. $\overline{\mu}_i$ is negative for 16 of the 22 physicians; $\overline{\mu}_i \geq 0$ for the remainder (Table A.3.1). Thus, weakly significanly more physicians choose quantities smaller than q_j^* ($p = 0.052$, binomial test, two-sided). Next we investigate whether underprovision is related to patient types.

Result 4. *Underprovision in CAP depends on patient types.*

Support: We again apply an order test and include those 19 subjects whose observed order comprises three different values (see Table A.3.2). Also in CAP, the order test reveals choices to be heavily dependent on patient types. We observe 0.158 average inversions. Thus the null hypothesis can be rejected at the 1% level. Analyzing the data in more detail shows that although patients are underserved on average, the number of physicians choosing q_j^* is larger than the number of physicians not choosing q_j^* for all patients of type 1 and 2. This is significant for four patients of type 2 (binomial test two-sided; see line I/CAP of Table A.3.3).Patients of type 3 are underserved in that the number of physicians choosing $q_j < q_j^*$ is larger than the number of physicians choosing q_j^*.This is weakly significant for one patient of type 3 (binomial test two-sided; see line I/CAP in Table A.3.3). Moreover, the level of underprovision $\overline{\nu}_j$ is highest for patient type 3 and lowest for patient type 2 (see Table A.3.4).

The maximum profit $\pi(\hat{q}_j)$ a physician can achieve in CAP is 12.00 Taler for all illnesses. Physicians' actual quantity decisions resulted in an average profit $\overline{\pi}(q_j)$ of 9.79 Taler (median 9.50 Taler, s.d. 1.52 Taler), i.e., 18% lower than $\overline{\pi}(\hat{q}_j)$. Average profits for each physician i vary between 7.84 and 11.48 Taler (see Table A.3.5). Testing over all patients, $\overline{\pi}(q_j)$ is highly significantly lower than $\overline{\pi}(\hat{q}_j)$ ($p = 0.000$, Wilcoxon signed ranks test, two-sided). How are profits affected by patient types in CAP? Table A.3.4 shows $\Delta\pi_{ij}$ averaged over all physicians. Highest deviations of 25 to 30% are found for patients of type 3 whereas lowest deviations of 7 to 11% occur for patients of type 2. Average profit deviations are 18.75% (8.71%) for patients of type 1 (2) and 27.67% for those of type 3.

The maximal average benefit a patient could gain in CAP, like in FFS, is 9.82 Taler if physicians always provided the patient optimal quantity. Actual average patient benefit is 8.56 Taler (median 9.75 Taler, s.d. 2.46 Taler), i.e., 13% lower than $\overline{B}(q_j^*)$. Further, average patient benefits determined by physician i vary between 2.73 and 9.82 Taler (see Table A.3.5).

To sum up, under CAP patients are underprovided in that physicians on average choose quantities of medical services smaller than the patient's optimal quantity. Provision of services and the deviation of profits from the profit maximum are strongly influenced by patient types, i.e., with increasing needs for services the levels of underprovision and profit deviations tend to increase. Also in CAP, physicians do not strive for the maximal profit. Patients received a benefit being on average 13% lower than the maximum benefit.

3.5.3 Comparison of behavior between FFS and CAP

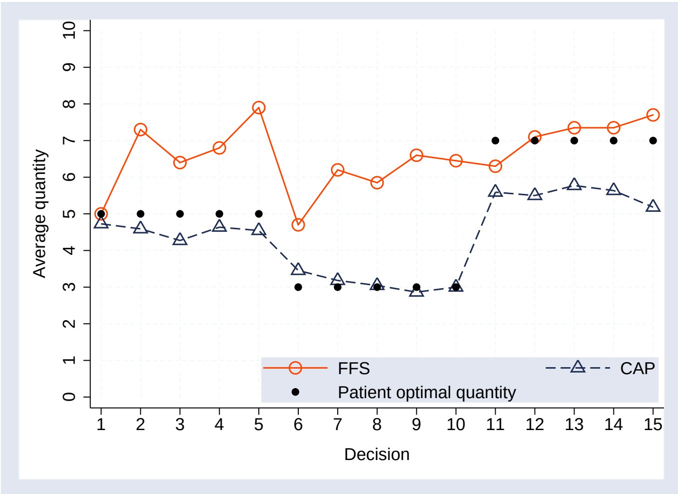
Our third research question is related to the consequences of *both* payment systems. We investigate differences in physicians' quantity choices across treatments and how patient types are affected. We compare physicians' profits, the provision of medical services, deviations from q_j^*, and patient benefit losses across payment systems for all patients and for patient types separately. The results above have already shown that physicians choose more medical services in FFS than in CAP. Thus, our next result implicitly follows from Results FFS1 and CAP1.

Result 5. *Patients are provided with more medical services in FFS than in CAP.*

SUPPORT: Evidence is provided by Figure 3.4 showing the average quantity of medical services per decision (patient) in both treatments. Not only do physicians in FFS on average provide 50% more services than in CAP (6.60 vs. 4.40; median: 7.00 vs. 5.00; s.d.: 1.85 vs. 1.64) but for each decision j, $\overline{q}_j^{FFS} > \overline{q}_j^{CAP}$. This is highly significant ($p = 0.0000$, Mann-Whitney U test, two-sided). The picture is similar when comparing individual decisions across treatments for each patient. Except for patients 1A and 3A, q_{ij}^{FFS} is significantly larger than q_{ij}^{CAP} ($p \leq 0.0010$, Mann-Whitney U test, two-sided; see line II of Table A.3.3). Thus, in FFS a significantly higher number of patients is provided with significantly more medical services compared to CAP ($p = 0.007$, binomial test, two-sided).

Physician's own profit $\pi(q_{ij})$ certainly is an important behavioral determinant in both treatments. As already mentioned, choosing $\hat{q}_j$ for all j in FFS would have yielded an average payoff $\overline{\pi}(\hat{q}_j)$ of 11.08 Taler. In CAP, the maximal profit is 12.00 Taler for all illnesses. They provided quantities of medical services such that their average profits are very similar in both treatments but about 17% lower than $\overline{\pi}(\hat{q}_j)$ (FFS: 9.17 Taler, CAP: 9.79 Taler). Average profits for each physician i vary between 6.53 and 10.93 Taler in FFS and between 7.84 and 11.48 Taler in CAP. In both payment systems, the average physician does not aim at the maximal achievable profit even though single physicians

Figure 3.4: Average quantity of medical services per patient



come very close to $\overline{\pi}(\hat{q}_j)$ (see Table A.3.5). To answer the question how profits are affected by patient types, we compare $\Delta\pi_{ij}$ across treatments. Except for patients 1C, 1E, 2C, and 3E, we find (weakly) significant differences between treatments.[15] For patients of type 2, $\Delta\overline{\pi}_{ij}^{FFS} > \Delta\overline{\pi}_{ij}^{CAP}$, for patients of type 3 the reverse holds.

We next compare the two payment systems with regard to how patients' health status is affected by physicians' choices. To this end, we first focus on the optimal treatment and deviations thereof. We then concentrate on the benefit losses patients suffer on average when some or all of them are not treated optimally.

Result 6. *Patient optimal quantities exert a stronger influence on physicians' behavior in CAP than in FFS.*

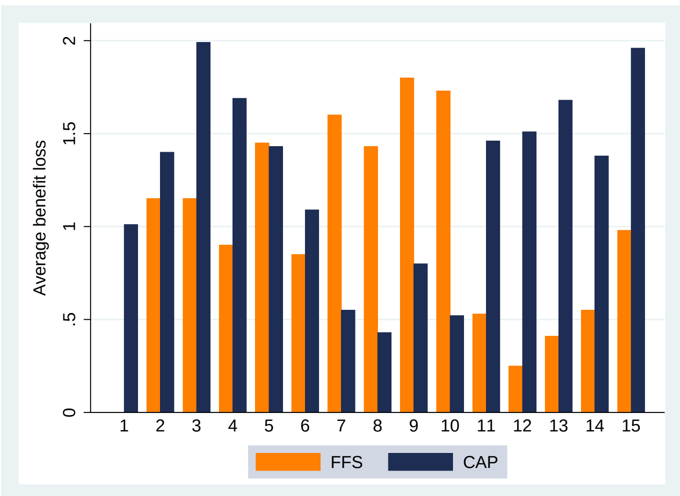
SUPPORT: Support comes from analyzing physicians' choices with regard to the patient optimal quantity. In CAP, the percentage of physicians choosing q_j^* per patient is significantly higher than in FFS ($p = 0.014$, Mann-Whitney U test, two-sided). If physicians deviate they tend towards opposite directions; a significantly larger share provides services larger than q_j^* in FFS compared to CAP ($p = 0.000$, Fisher exact test, two-sided). In FFS, $\overline{\mu}_i > 0$ except for physicians $i = 3, 4, 17$; in CAP, $\overline{\mu}_i \leq 0$ except for physicians $i = 4, 19$ (see Table A.3.1). Analyzing patient types separately, we find all patients of type 2 in CAP to get a better treatment in that significantly more physicians per patient choose q_j^* compared to FFS ($p \leq 0.011$, Fisher exact test; see line III in Table A.3.3). The same applies to

[15]For type 1: $p \leq 0.059$; for type 2: $p \leq 0.018$; for type 3: $p = 0.000$, all Mann-Whitney U test, two-sided; see line IV of Table A.3.3.

all patients of type 1 except for patient 1A[16] ($p \leq 0.009$, Fisher exact test). Evidence is mixed for patients of type 3. We find no significant difference for patients 3A, 3C, 3E. For patients 3B and 3D physicians choose q_j^* significantly more often in FFS than in CAP (see line III of Table A.3.3). In both treatments, the benefit maximum for patients of type 3 ($B_3(q_j^*)$) is 9.45 Taler. $B_1(q_j^*) = B_2(q_j^*) = 10$ Taler resulting in $\overline{B}(q_j^*) = 9.82$ in FFS *and* in CAP. Our experimental physicians actually provide quantities of medical services such that the average patient benefit $\overline{B}(q_j)$ was slightly larger in FFS (8.83 Taler) than in CAP (8.56 Taler) and around 10% smaller than $\overline{B}(q_j^*)$. These numbers seem to suggest that nearly no differences between payment systems exist. Yet, the picture is different when having a closer look at the data. Focusing on single patients and their health status we, like Newhouse and the Insurance Experiment Group (1993), find patients to be affected differently by the mode of payment. Moreover, average patient benefits vary between 7.52 and 9.82 Taler in FFS and between 2.73 and 9.82 Taler in CAP (see Table A.3.5).

Whenever a physician i deviates from choosing q_j^*—when patients are either under- or overprovided—patients suffer a benefit loss, as they do not obtain the treatment rendering the maximum benefit. Let $\psi(q_{ji}) := |B(q_{ij}) - B(q_j^*)|$.

Figure 3.5: Average benefit loss per patient



Result 7. *Benefit losses per patient depend on patient types and differ across treatments.*

Support: Figure 3.5 contrasts the average benefit loss per patient across treatments. For 10 of the 15 patients, $\overline{\psi}(q_j)^{CAP} > \overline{\psi}(q_j)^{FFS}$. The benefits loss in FFS is larger for the

[16] Here, physicians in FFS make significantly more q_j^*-choices ($p = 0.006$, Fisher exact test; see line III in Table A.3.3).

remaining patients (see also Table A.3.4). Benefit losses differ significantly for all patients of type 2 ($p \leq 0.027$, Mann-Whitney U test, two-sided, see line IV of Table A.3.3). Losses are larger in FFS for patients 2B, ..., 2E; the reverse holds for patient 2A. For 9 of the 10 patients of types 1 and 3, benefit losses in CAP are larger than in FFS. Differences are only significant for two patients of type 1 and 3 each.[17]

Result 7 suggests that for patients in need of a small quantity of medical services like patients of type 2, a smaller benefit loss results when physicians are paid by CAP. Patients in need of a larger quantity of medical services, like patients of types 1 and 3, incur a smaller loss under FFS.

In short, the cross-treatment comparison indicates that physicians' choices of medical services are influenced by incentives embeded in FFS and CAP payment systems. Physicians under FFS choose significantly more medical services than those under CAP. Consequently, the mode of payment does affect patients' health status. In particular, patients of type 1 and 2 are treated more optimally under CAP than under FFS and the patient benefit loss is significantly smaller in the former payment system for all but one of patients of type 2.

3.5.4 Tradeoffs and Pareto efficiency

In this section, we are concerned with the tradeoff between own profit and patient benefit a physician encounters when making a quantity decision. In particular, we investigate the Pareto efficiency of physician's choices.

In general, Pareto efficiency means that an allocation X is Pareto preferred to another allocation Y if at least one person is better off and no one is worse off with X than with Y. Besides its importance in general economic theory, the concept of Pareto efficiency also plays a prominent role in health economics (e.g., Iversen 1993, De Jaegher and Jegers 2000 and Olivella and Vera-Hernandez 2007). In the context of our experiment a situation is said to be Pareto efficient, if no unanimous move to another allocation of profit and patient benefit is possible. That means, a Pareto-efficient (PE) choice involves that changing q can neither make the physician better off without inducing a benefit loss to the patient nor make the patient better off without foregoing own profit. Pareto-inefficient (PIE) choices do not involve a benefit/profit tradeoff as changing q can increase *both* a physician's own profit and the patient benefit; they are dominated by Pareto-efficient choices.

[17] 1A (where no losses occur in FFS): $p = 0.009$; 1E: $p = 0.062$; 3B: $p = 0.002$; 3C: $p = 0.050$, all Mann-Whitney U test, two-sided, see line IV of Table A.3.3.

Pareto-efficient decision options exist for each patient in both treatments. The number of PE benefit/profit pairs differs according to illnesses (in FFS only) and patient types. In FFS, physicians can choose between one and eight PE decisions per patient. In CAP, there are either four (patient type 2), six (patient type 1) or eight (patient type 3) PE pairs. PE choices are positioned on the upper right line in Figures A.3.1 and A.3.2, the Pareto frontier, whereas PIE decisions are those below the Pareto frontier.

It is remarkable that 597 of the 660 choices are Pareto-efficient. Thus, 95% of all physicians' actual choices both in FFS and CAP involve a tradeoff between physician's own profit and patient benefit. Pareto efficiency guides all the decisions by 13 of the 20 physicians (65%) in FFS and by 15 of the 22 physicians (68%) in CAP. The remaining choices entail up to 4 (9) PIE decisions per physician in FFS (CAP). Hence, not only has the majority of physicians Pareto efficiency as their only target but also the remaining physicians behave accordingly with the vast majority of their quantity decisions.

To further characterize physicians' choices we subdivide the set of PE decisions into categories capturing variables of economic importance and medical ethics: own profit maximum, patient benefit optimum, social optimum.

- PROMAX comprises choosing $\hat{q_j}$, the profit-maximizing quantity of medical services. The corresponding benefit/profit pair is $(B(\hat{q_j}), \pi(\hat{q_j}))$.
- PATMAX consists of q_j^*-choices maximizing the patient's benefit. $(B(q_j^*), \pi(q_j^*))$ is the resultant benefit/profit pair.
  PROMAX and PATMAX are the two boundary points of the Pareto frontier (see Figures A.3.1 and A.3.2).
- SOCOPT is suggested by a welfare economics perspective and contains the socially optimal choices, i.e., decisions where $(\pi(q_j) + B(q_j))$ is maximal.
  Note that patients exist where SOCOPT coincides with PROMAX and/or PATMAX (Table A.3.1). Only those decisions are assigned to SOCOPT that are not yet covered by the two previous categories.[18]
- PAROTH is a residual category comprising the remaining benefit/profit pairs on the Pareto frontier not included in any of the other three categories.

[18]We decided on this assignment as subjects choices may not be motivated by the social optimum in the first place for the following reason. Finding q^{soc} seems not straightforward as participants first have to calculate $\pi(q_j) + B(q_j)$ and then they have to determine the maximum. Selecting $\hat{q}$ or q^* is much more obvious given the information on the decision screens (see the instructions in Appendix).

In FFS, 16% of all physician's Pareto-efficient choices are assigned to PROMAX, 34% to PATMAX, 16% to SOCOPT and 34% to PAROTH.[19] In CAP, only 2% of physicians' choices are attributed to PROMAX, 66% to PATMAX, 6% to SOCOPT and 26% are covered by PAROTH. [20]

Comparing both payment systems, a much lower percentage of decisions in CAP is motivated by $\pi(\hat{q_j})$ probably because choosing $\hat{q_j}^{CAP}$ entails *no* provision of services to the patient. Such behavior would be a severe violation of the professional code of medical ethics. Noticeably, two thirds of all Pareto-efficient decisions in CAP involve $B(q_j^*)$ versus only one third in FFS. This may be due to the fact that choosing q^{*CAP} implies a lower own-profit reduction than in FFS where the physician on average forgoes 39.6% of her maximally achievable profit vs. only 23.3% in CAP. The social optimum plays no role in CAP possibly because q^{soc} coincides with q^* for all 10 patients of types 1 and 2.
In short, our analysis provided compelling results. First, nearly all physicians' decisions are Pareto efficient. Second, the vast majority of these choices (66% in FFS and 74% in CAP) can be explained by motives based on variables of economic and ethical importance.

3.6 Conclusion

The paper introduces a controlled laboratory experiment to test for the influence of payment systems on physicians' provision behavior. By assigning the monetary equivalent of the patient benefit to treating actual patients we substituted the 'abstract' patients in our experiment with real ones.

Our results are in line with the theoretical literature (e.g., Ellis and McGuire 1986) and add further evidence to previous findings in the field. Patients are overserved in FFS in that experimental physicians on average choose quantities of medical services larger than the patient's optimal quantity. Provision is dependent on patient types as is the deviation of profits from the profit maximum. The cross-treatment comparison indicates that physicians' choices are influenced by incentives ebedded in the two payment systems. Physicians in FFS provide more medical services than those in CAP do. Like Newhouse and the Insurance Experiment Group (1993), we found the mode of payment to affect patients' health status. Patients in need of a low level of medical services are better off under CAP, whereas patients with a high need of medical services gain more health benefit

[19] When calculating the percentages, $j = 1$ is neglected because here categories PROMAX, PATMAX and SOCOPT coincide and we cannot even distinguish whether $q = 5$ was motivated by $\hat{q}_j$ or by q_j^*.

[20] A detailed overview on relative frequencies per category is provided in Table A.3.1.

when physicians are paid by FFS. How these gains and losses are to be weighed against each other is a matter of political decision, however.

In both remuneration systems, financial incentives are not the only motivation for physicians' quantity decisions, though. As the patient benefit is of considerable importance, patients received a substantial benefit the financial equivalent of which allowed to treat nine real patients by ophthalmic surgery.

Experiments in health economics might serve as a 'wind tunnel' or 'test bed' before institutional changes are implemented during a health care reform. Even though an experiment always simplifies a physician's decision task when caring for a patient it, at the same time, allows to separating behavioral determinants. While simplifications give rise to caution when extrapolating the results, they also suggest the lines for further experimental research like introducing uncertainty about the impact of medical treatments and patients' health status, patients' demand effects and monitoring mechanisms.

Chapter 4

Do prospective physicians behave differently?

4.1 Introduction

Other-regarding motivations are a fundamental determinant of public service provision and are, thus, of considerable importance when designing robust incentives structures in the public sector (Le Grand 2003). Most public service providers, Le Grand argues, derive a great personal satisfaction from helping other people. In this respect, public service providers behave like 'act-relevant knights' who help needful people, e.g., physicians providing medical services to their patients.[1] At the same time their motivation to act knightly (altruistically) also seems to depend upon the degree of personal sacrifice associated with the act.

One easily agrees with Le Grand who states that payment incentives in the public sector—and in particular in the health care sector—need to be such that 'knightly' behavior is fostered. An incentive scheme designed to motivate the 'standard neoclassical worker' (assumed to be without an other-regarding motivation) might be unsuitable for this purpose. Indeed, recent theoretical literature departs from the assumption of the neoclassical worker by acknowledging the importance of motivations of public service providers (e.g., Francois 2000, Besley and Ghatak 2005, Delfgaauw 2007, Delfgaauw and Dur 2008). Despite the acknowledged importance of other-regarding motivations in the theoretical literature, little is known about their relationship to payment incentives empirically.

In our study we explicitly explore this relationship with the help of a controlled lab-

[1]Notice that in several theoretical models describing physician-patient interaction an 'ethical' argument is used to represent the physician's regard not only for professional codes of conduct but also for a 'knightly' or altruistic motivation to care for patients' welfare (see, e.g., Woodward and Warren-Boulton 1984, Ellis and McGuire, 1986 and McGuire, 2000).

oratory experiment in a medical decision-making context (equivalent to Hennig-Schmidt et al., 2009). In particular, we compare behavior of prospective physicians, i.e., medical students (Meds), assumed to exhibit other-regarding motivations and a 'standard' subject pool in economic experiments, i.e., students from various fields of study other than medicine (Non-Meds) assumed to be without such motivations. Decisions are incentivized by fee-for-service (FFS) and capitation (CAP)—the two prevalent payment methods for physicians in the public health care sector (Newhouse 2002). The experimental economic method allows us to clearly identify possible knightly behavior and to quantify individual sacrifices. A questionnaire succeeding the experiment collects information about subjects' motivations.

Our study also contributes to the experimental economic literature investigating behavioral influences of subject pools. Differences in behavior across various subject pools have already been highlighted in several experiments; for summaries see Ball and Cech (1996) and Croson and Gneezy (2009). For example, substantial evidence suggests that economics students are less cooperative, altruistic or trusting than other students in a variety of experimental contexts such as public goods (Marwell and Ames 1981, Cadsby and Maynes 1998) and bargaining (Carter and Irons 1991, Kahneman et al. 1986). To control for possible effects due to more selfish behavior of economics students the number of economics students among the Non-Meds is well balanced in our experiment.

Recent laboratory experiments in health economics indicate the increasing importance of the experimental method in this area. Fan et al. (1998) study alternative methods for controlling the cost of physician services under global budgeting. The experiment of Hennig-Schmidt et al. (2009) analyzes incentives from different payment systems on physicians' provision behavior. Ahlert et al. (2008) find less selfishness of medical students compared to economics students in a medically framed decision task. The experimental studies by Schram and Sonnemans (2008) and Lévy-Garboua et al. (2008) analyze issues dealing with health care funding and health insurance choice. Recent experimental literature on the economics of credence goods is also related to our study, in the sense, that a physician-patient setting might be regarded as a special case of a expert-laymen relationship (Dulleck et al. forthcoming).[2] For experimental investigations in a medical decision-making context it is thus of considerable importance to know whether there are differences in behavior across Meds and Non-Meds, as the choice of the subject pool could

[2]For a comprehensive overview about the economics of credence goods see Dulleck and Kerschbamer (2006).

drive experimental results.

Our main finding is that Meds and Non-Meds differs substantially in their provision of medical services. Meds' decisions are more strongly directed towards the patients' benefit leading to a lower tendency to overprovide (underprovide) patients under FFS (CAP) than observed for Non-Meds. Le Grand (2003) assumes that subjects' motivation to behave like 'act-relevant knights' seems to depend positively upon the degree of personal sacrifice associated with the decision. Our experimental results provide evidence for this assumption as the behavioral differences can be explained by Meds' higher willingness to sacrifice own profit to increase a patient's benefit compared to Non-Meds. Economists are less willing to sacrifice own profit than non-economists when paid by CAP and, thus, drive behavioral differences between Meds and Non-Meds under this payment method. Under FFS, however, proportional sacrifices between economists and non-economists do not differ significantly.

The remainder of the chapter proceeds as follows. Section 4.2 summarizes the experimental design and procedure. The results on subjects' behavior, willingness to sacrifice and stated motivations are presented in Section 4.3. Section 4.4 concludes.

4.2 Experimental design and procedure

4.2.1 Design and parameters

In our experiment we study behavior of prospective physicians (Meds), i.e., medical students expected to become physicians in the future, and non-medical students (Non-Meds) in a medical decision-making context. Subjects decide in the role of physicians on the provision of medical services either incentivized by fee-for-service (FFS) or capitation (CAP) payments analogous to the study by Hennig-Schmidt et al. (2009). When paid by fee-for-service (treatments Med-FFS, Non-Med-FFS), subjects receive a monetary reward for each unit of medical services provided. Under capitation (treatments Med-CAP, Non-Med-CAP) subjects receive a lump-sum payment per patient independent of the number of medical services. Table 4.1 provides an overview of our 2×2 design.

Non-Meds are recruited from various fields of study. The number of economists is well balanced as 10 economists participate in treatments Non-Med-FFS and Non-Med-CAP. Any subject participates in only one treatment. Each subject i makes 15 individual decisions j on the quantity of medical services, i.e., $q \in \{0, 1, \ldots, 10\}$, for three patient types $(1, 2, 3)$ with five abstract illnesses A, B, C, D, E. Patient types differ in benefits gained from medical services ($B_1(q)$, $B_2(q)$, $B_3(q)$). The patient benefit is measured in monetary

Table 4.1: Experimental treatments

Treatment	Subject pool	Payment condition	Number of subjects
Med-FFS	Medical students	Fee-for-service	20
Med-CAP	Medical students	Capitation	22
Non-med-FFS	Non-medical students	Fee-for-service	22
Non-med-CAP	Non-medical students	Capitation	22

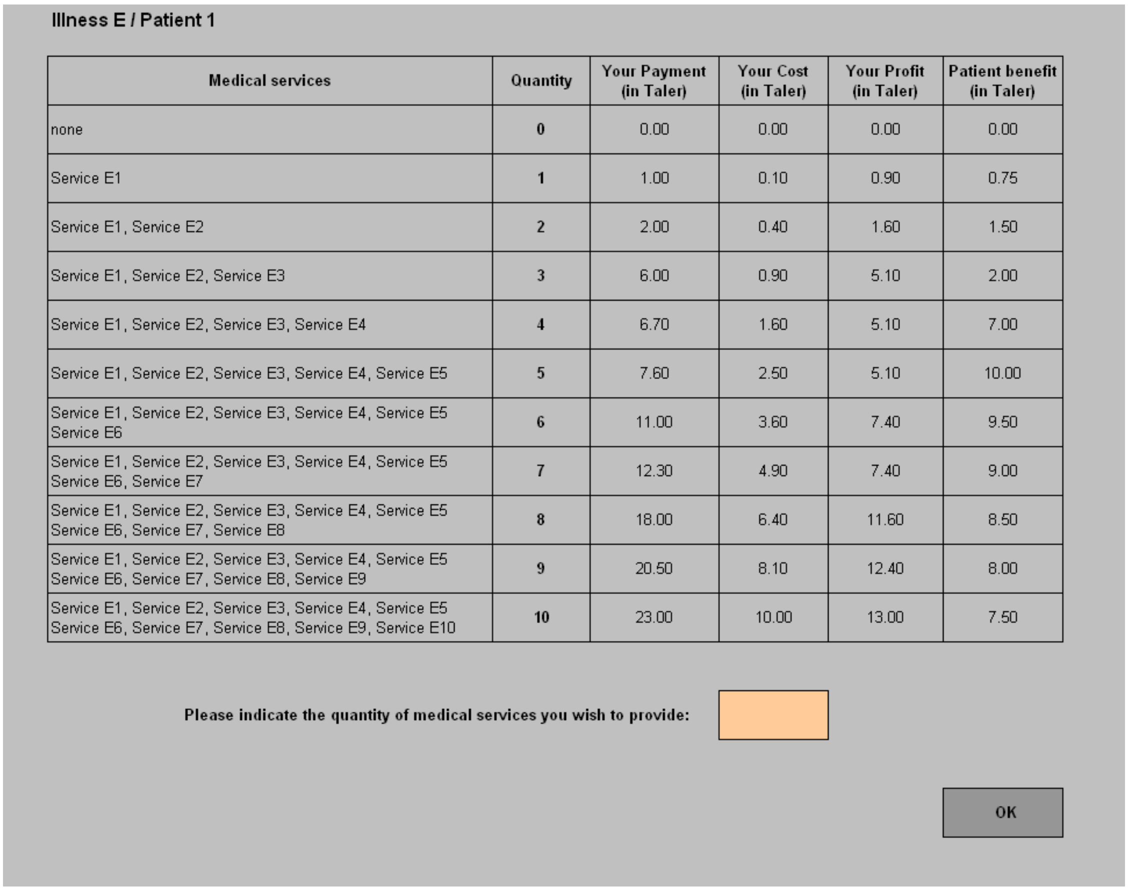
terms. The combination of patient type and illness characterizes a single patient, i.e., 1A, 1B, 1C, ..., 3D, 3E. By choosing a quantity for each patient subjects simultaneously determine their own profit ($\pi(q)$) and the patient benefit. The patient is assumed to be passive and fully insured accepting each level of medical service provided by the physician.

As only abstract patients occur in our experiment, we provided the experimental physicians with an incentive to care for real patients outside the lab. Subjects are informed that the aggregated patient benefit resulting from their decisions is transferred to a charity caring for real patients. To verify that the money is actually transferred we apply a procedure similar to Eckel and Grossman (1996).

An example of a decision situation in treatments Med-FFS and Non-Med-FFS is provided in Figure 4.1. Subjects determine the quantity of medical services they desire to provide for patient 1E ($j = 5$). They are informed about their payment, costs and profit as well as the patient's benefit for each quantity from 0 to 10. All monetary amounts are depicted in Taler, our experimental currency (1 Taler = EUR 0.05). The first two columns show abstract descriptions of medical services and the corresponding quantities. The third column indicates subjects' payment increasing in the quantity of medical services. Notice that in treatments Med-CAP and Non-Med-CAP the payment remains the same for all quantities as subjects receive a lump-sum payment. Costs of medical services, being constant across all treatments, are shown in the fourth column. Physician's profit (payment minus costs) is given in the fifth column and the patient's benefit in the final column. There is a tradeoff between 'own' maximum profit and patient optimal benefit. In this example, a quantity of 5 medical services renders the patient benefit to be optimal ($B(5) = 10.00$, $\pi(5) = 5.10$) whereas the profit maximizing quantity is 10 ($B(10) = 7.50$, $\pi(10) = 13.00$). Quantity choices between 5 and 10 are Pareto optimal, in that, subjects cannot increase the patient's benefit without foregoing own profit and vice versa.[3]

[3]For example, choosing 8 rather than 7 medical services would provide patient 1E with a slightly larger benefit (8.50 instead of 9.00) while the physician's profit would increase substantially from 7.40 to 11.60.

Figure 4.1: Example of a decision screen in Med-FFS and Non-Med-FFS



Illness E / Patient 1

Medical services	Quantity	Your Payment (in Taler)	Your Cost (in Taler)	Your Profit (in Taler)	Patient benefit (in Taler)
none	0	0.00	0.00	0.00	0.00
Service E1	1	1.00	0.10	0.90	0.75
Service E1, Service E2	2	2.00	0.40	1.60	1.50
Service E1, Service E2, Service E3	3	6.00	0.90	5.10	2.00
Service E1, Service E2, Service E3, Service E4	4	6.70	1.60	5.10	7.00
Service E1, Service E2, Service E3, Service E4, Service E5	5	7.60	2.50	5.10	10.00
Service E1, Service E2, Service E3, Service E4, Service E5 Service E6	6	11.00	3.60	7.40	9.50
Service E1, Service E2, Service E3, Service E4, Service E5 Service E6, Service E7	7	12.30	4.90	7.40	9.00
Service E1, Service E2, Service E3, Service E4, Service E5 Service E6, Service E7, Service E8	8	18.00	6.40	11.60	8.50
Service E1, Service E2, Service E3, Service E4, Service E5 Service E6, Service E7, Service E8, Service E9	9	20.50	8.10	12.40	8.00
Service E1, Service E2, Service E3, Service E4, Service E5 Service E6, Service E7, Service E8, Service E9, Service E10	10	23.00	10.00	13.00	7.50

Please indicate the quantity of medical services you wish to provide:

OK

Next we briefly describe the experimental parameters. Under FFS (treatments Med-FFS and Non-Med-FFS) subjects receive a fee for each unit of medical services provided. The payment differs across illnesses, i.e., $R_A(q), R_B(q), \ldots, R_E(q)$. In treatments Med-CAP and Non-Med-CAP, subjects are paid a lump-sum of 12 Taler per patient (see panel I of Table 4.2).

To take a heterogeneous patient population into account being observed in the real world, e.g., with regard to a patient's health status or severity of illness, we consider three patient types differing in their benefit gained from medical services. A common characteristic of all patient types is a global optimum on the quantity interval from 0 to 10 (see panel IV of Table 4.2). Intuitively, it might be interpreted as the medically agreed best number of medical services. Taking the patient optimal quantity q^* as the benchmark, we can identify overprovision and underprovision of medical services. Patient types vary with respect to the optimal quantity: $q_j^* = 5$ for patient type 1 ($j \in [1, 5]$), $q_j^* = 3$ for patient type 2 ($j \in [6, 10]$) and $q_j^* = 7$ for patient type 3 ($j \in [11, 15]$).

Further parameters relevant for physicians' decisions are costs $c(q)$ and profit $\pi(q)$ shown in panels II and III of Table 4.2. Like real doctors, physicians in the experiment have to bear costs depending on the quantity of medical services they choose. Costs are assumed to be $c(q) = \frac{1}{10} \cdot q^2$ in all treatments. Profit varies across illnesses in treatments Med-FFS and Non-Med-FFS because remuneration differs for illnesses, and costs are constant. In Med-CAP and Non-Med-CAP, profit does not vary with illnesses. In FFS, the profit maximizing quantity $\hat{q}_j$ is 10 for all patients, except for patients 1A, 2A and 3A where $\hat{q}_j = 5$. Notice that for patient 1A patient optimal and profit maximizing quantity coincide, as $\hat{q}_1 = q_1^* = 5$. Under CAP, a profit maximizing subject would not provide any medical service at all, i.e., $\hat{q}_j = 0 \ \ \forall j \in [1, 15]$.

4.2.2 Experimental protocol

The computerized experiment programmed with z-Tree (Fischbacher 2007) was conducted at Bonn$E$con$L$ab, the Laboratory for Experimental Economics at the University of Bonn. Overall 86 students recruited by the online recruiting system ORSEE (Greiner 2004) participated in the experimental sessions: 20 (22) Meds in FFS (CAP) and 22 Non-Meds in both FFS and CAP.

Upon arrival, subjects were randomly allocated to the cubicles. Then, the experimenter read the instructions aloud. Subjects was given plenty of time for clarifying questions which were asked and answered in private. To check for subjects' understanding of the decision task they had to answer three test questions. The experiment did not start unless all participants had answered the test questions correctly.

The subjects then made their 15 individual quantity choices in complete anonymity. The order of decision screens was predetermined and kept constant across treatments. Having made their choices, subjects were asked to fill in a computerized questionnaire explaining their motivations and the factors having influenced their decisions. Finally, the monitor was assigned randomly. After the experiment, subjects were paid in private according to their choices. At last, a randomly selected monitor verified that a check on the benefits of all patients was written and sealed in an envelope addressed to the *Christoffel Blindenmission*. The monitor and experimenter then walked together to the nearest mailbox and deposited the envelope. The monitor was paid an additional 4.00 EUR (see also the instructions in Appendix A.3.2).

Sessions lasted for about 30 to 40 minutes. Meds (Non-Meds), on average, earned 6.88 EUR (7.71 EUR) in FFS and 7.42 EUR (7.80 EUR) in CAP. In treatments with Meds

Table 4.2: Experimental parameters

	Payment	Var.	Quantity (q) 0	1	2	3	4	5	6	7	8	9	10
I	FFS	$R_A(q)$	0.00	1.70	3.40	5.10	5.80	10.50	11.00	12.10	13.50	14.90	16.60
		$R_B(q)$	0.00	1.00	2.40	3.50	8.00	8.40	9.40	16.00	18.00	20.00	22.50
		$R_C(q)$	0.00	1.80	3.60	5.40	7.20	9.00	10.80	12.60	14.40	16.20	18.30
		$R_D(q)$	0.00	2.00	4.00	6.00	8.00	8.20	15.00	16.90	18.90	21.30	23.60
		$R_E(q)$	0.00	1.00	2.00	6.00	6.70	7.60	11.00	12.30	18.00	20.50	23.00
	CAP	$R(q)$	12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00
II	FFS, CAP	$c(q)$	0.10	0.20	0.40	0.90	1.60	2.50	3.60	4.90	6.40	8.10	10.00
III	FFS	$\pi_A(q)$	0.00	1.60	3.00	4.20	4.20	8.00	7.40	7.20	7.10	6.80	6.60
		$\pi_B(q)$	0.00	0.90	2.00	2.60	6.40	5.90	5.80	11.10	11.60	11.90	12.50
		$\pi_C(q)$	0.00	1.70	3.20	4.50	5.60	6.50	7.20	7.70	8.00	8.10	8.30
		$\pi_D(q)$	0.00	1.90	3.60	5.10	6.40	5.50	11.40	12.00	12.50	13.20	13.60
		$\pi_E(q)$	0.00	0.90	1.60	5.10	5.10	5.10	7.40	7.40	11.60	12.40	13.00
	CAP	$\pi(q)$	12.00	11.90	11.60	11.10	10.40	9.50	8.40	7.10	5.60	3.90	2.00
IV	FFS, CAP	$B_1(q)$	0.00	0.75	1.50	2.00	7.00	10.00	9.50	9.00	8.50	8.00	7.50
		$B_2(q)$	0.00	1.00	1.50	10.00	9.50	9.00	8.50	8.00	7.50	7.00	6.50
		$B_3(q)$	0.00	0.75	2.20	4.05	6.00	7.75	9.00	9.45	8.80	6.75	3.00

This table shows all experimental parameters. $R(q)$ denotes physicians' remuneration. Under FFS, $R(q)$ varies with illnesses A, . . . , E and increases in q, whereas, under CAP, $R(q)$ remains constant. The costs for providing medical services $c(q)$ increase in q and are equivalent under FFS and CAP. Physicians' profit $\pi(q)$ is equal to remuneration $R(q)$ minus costs $c(q)$. $B_1(q), B_3(q), B_3(q)$ denote the patient benefit for the three patient types being equivalent in FFS and CAP.

(Non-Meds) a total of 273.68 EUR (241.80 EUR) was transferred to the charity, 6.62 EUR (6.05 EUR) per participant in FFS and 6.42 EUR (4.94 EUR) in CAP. The money was assigned to support surgical treatments of cataract patients in a hospital in Masvingo (Zimbabwe) staffed by ophthalmologists of the *Christoffel Blindenmission*. Subjects were not informed about the money being assigned to a developing country (see instructions in Appendix A.3.2) because we wanted to exclude motives of compassion. Average costs for a cataract surgery is about 30.00 EUR. Thus, the aggregated patient benefit from Meds (Non-Meds) allowed to treat nine (eight) patients.

4.3 Results

We report the results in three parts. First, we explore the behavioral data for potential differences across subject pools. Second, we analyze subjects' willingness to sacrifice own profit to increase the patient benefit. Here, we also investigate behavior of economists and non-economists among the Non-Meds separately. Finally, we compare our findings with questionnaire data on subjects' motivations.

4.3.1 Provision behavior

When comparing subjects' aggregate behavior we observe substantial differences across subject pools in both payment conditions.[4] Under FFS, Non-Meds choose substantially larger quantities of medical services than Meds. The average quantity provided by Non-Meds (7.74, s.d. 1.90) is about 15% larger than for Meds (6.60, s.d. 1.85). The difference in CAP is even more pronounced as the average quantity provided by Non-Meds is 20% lower than for Meds (3.52, s.d. 1.89, versus 4.40, s.d. 1.64).

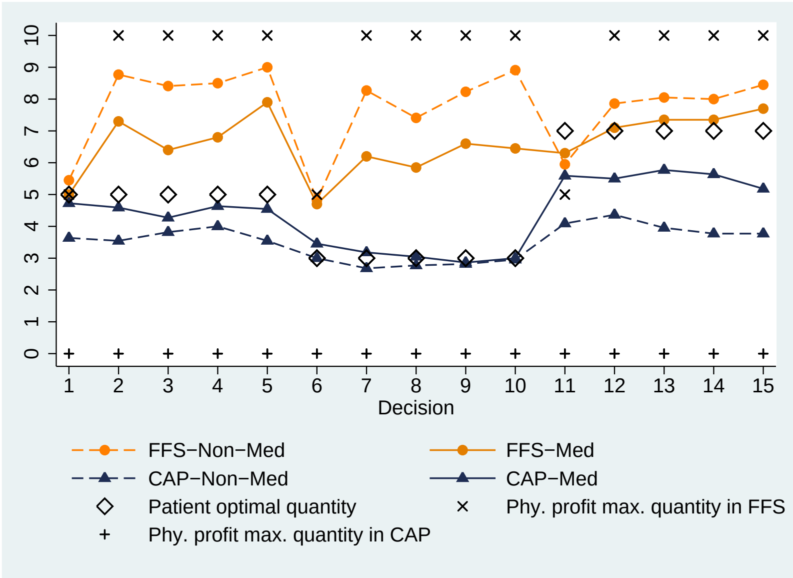
What do observed behavioral differences across Meds and Non-Meds imply for patients? A comparison of subjects' average quantities per patient in FFS yields the tendency to overserve patients to be larger for Non-Meds than Meds (see Figure 4.2). The difference across subject pools is highly significant ($p = 0.0034$, Mann-Whitney U-Test, two-sided).[5] Moreover, Figure 4.2 shows a stronger tendency towards underprovision of patients for Non-Meds compared to Meds in CAP. The average quantities provided in Med-CAP and Non-Med-CAP differ significantly ($p = 0.0144$).

We find that subjects' decisions are quite substantially influenced by the patient's optimal benefit, yet Meds significantly more so than Non-Meds. In 36.67% (61.52%) of

[4]Notice that data for treatments Med-FFS and Med-CAP are taken from Hennig-Schmidt et al. (2009).

[5]In the following all tests are two-sided Mann-Whitney U-Tests if not indicated differently.

Figure 4.2: Average quantities per decision across treatments



This figure shows average quantities of medical services for each decision (patient) for all treatments. Moreover, it shows the patient optimal quantity q^* and the profit maximizing quantity $\hat{q}$ for both payment conditions.

their choices Meds opt for the patient optimal quantity q^* under FFS (CAP). The average frequency Meds choose q^* is 5.50 (9.22) in FFS (CAP). Further, 10% (14%) of Meds in treatments Med-FFS (Med-CAP) *always* choose q^* *no* Med has always chosen the profit maximizing quantity $\hat{q}$. Data for Non-Meds reveal a different picture, as the tendency towards treating a patient optimally is less pronounced among Non-Meds. Here, 21.21% (38.18%) of choices are q^* in treatment Non-Med-FFS (Non-Med-CAP). Non-Meds choose, on average, only about 3 (6) times q^* under FFS (CAP). In Non-Med-FFS, even 9% of the subjects *always* choose $\hat{q}$. More concisely, comparing the number of individual q^*-choices reveals significant differences across subject pools under both payment conditions (FFS: $p = 0.0209$, CAP: $p = 0.0213$). In short, we state the following result.

Result 1. *Non-Meds provide significantly more (less) medical services under FFS (CAP) than Meds, thus Non-Meds overserve (underserve) patients to a higher extent than Meds. Moreover, Non-Meds choose the patient optimal quantity significantly less often than Meds.*

4.3.2 Willingness to sacrifice own profit

We now turn to the analysis of subjects' willingness to sacrifice own profit in order to increase a patient's benefit. Our approach is similar to Selten and Ockenfels (1998) who analyze proportional sacrifices in the the solidarity game, a variant of the dictator game. In the context of public service provision, Le Grand (2003) describes a personal sacrifice as a prerequisite for act-knightly behavior.

Our experimental design allows to analyze subjects' individual sacrifices. Recall that increasing the patient's benefit comes at the cost of sacrificing own profit for the subjects. In particular, subject i faces a tradeoff between achieving his/her maximum profit $\hat{\pi}_j$ and the patient's optimal benefit simultaneously for each decision j.[6] By $\rho_{ij} = (\pi_{ij} - \hat{\pi}_j)/(\pi^*_j - \hat{\pi}_j)$, where $\pi^*_j \equiv \pi(q^*_j)$, we measure actual sacrifices proportional to the maximal possible sacrifice. In particular, ρ_{ij} normalizes subjects' willingness to sacrifice on a 0-to-1 scale. When $\rho_{ij} = 1$, subject i chooses the patient optimal quantity q^*_j in decision j and is, thus, willing to sacrifice the profit necessary to provide the best medical treatment from a patient's perspective. On the contrary, when $\rho_{ij} = 0$ subject i chooses her own profit maximizing quantity $\hat{q}_j$ and is not willing to forego own profit to increase the patient's benefit. At $\rho_{ij} = \frac{1}{2}$, for example, $\pi^*_j - \hat{\pi}_j$ is twice as big as the actual made sacrifice. We say the higher ρ_{ij} the more a physician in the experiment is caring for a patient. See Figures A.4.1 to A.4.4 for ρ_{ij} in each of the four treatments. We restrict our analysis to the 94% of choices that are Pareto efficient.

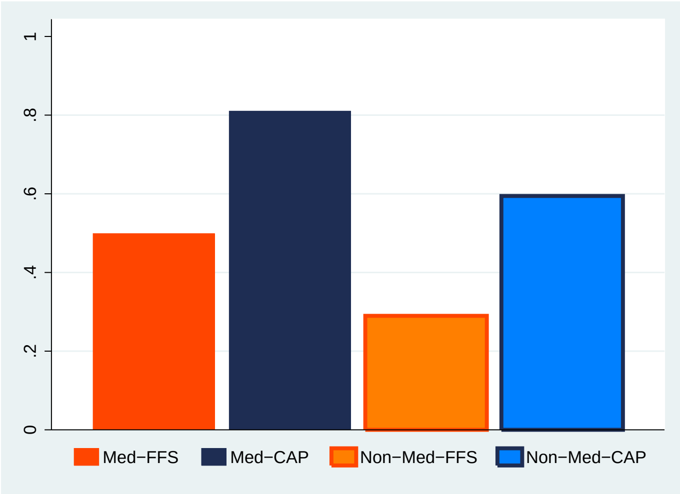
Overall the proportional sacrifices of Meds are 36.22% higher than of Non-Meds ($\overline{\rho}^{Med} = 0.6664$, s.d. 0.3824; $\overline{\rho}^{Non-Med} = 0.4250$, s.d. 0.4036). This difference is highly significant ($p = 0.0018$).

When differentiating between payment systems we also find significant differences across subject pools. As Figure 4.3 shows, Meds are more willing to sacrifice own profit than Non-Meds in both payment conditions. The proportional sacrifice under FFS is, on average, higher for Meds (0.5008, s.d. 0.4058) than for Non-Meds (0.2466, s.d. 0.3367). Under CAP we observe the same behavioral tendency. Here, the average proportional sacrifice of Meds is 0.8107 (s.d. 0.2934) and of Non-Meds 0.5949 (s.d. 0.3891), respectively. Both differences are significant (FFS: $p = 0.0095$, CAP: $p = 0.0218$). Comparing subjects' proportional sacrifices across payment systems yields that both Meds and Non-Meds are willing to forego, although at different levels, substantially more profit in CAP than in FFS ($p = 0.0000$). In short, the result is as follows.

Result 2. *Meds are less driven by profit maximizing objectives compared to Non-Meds as they exhibit a higher willingness to sacrifice own profit to increase patients' welfare. Put*

[6] We exclude decision $j = 1$ under FFS as no tradeoff is involved.

Figure 4.3: Average proportional sacrifices by subject pool and payment system



differently, Meds can be said to be more caring for their patients than Non-Meds.

In the following we investigate whether economists are less willing to sacrifice own profit and, thus, drive the lower willingness to sacrifice own profit of Non-Meds. Substantial evidence suggests that economics students are more selfish in a variety of experimental tasks and contexts (see Ball and Cech 1996). In a medically framed allocation task Ahlert et al. (2008) observe economists to behave more selfish than medical students. Do we also find evidence for a more selfish behavior of economists compared to Non-economists and Meds in the medical decision-making context of our experiment? Aggregated over both payment systems, we find that economics students are, on average, less caring for their patients ($\overline{\rho}^{Econ} = 0.3618$, s.d. 0.2037) than non-economist ($\overline{\rho}^{Non-Econ} = 0.4822$, s.d. 0.3779). This indicates a strong tendency towards a more selfish behavior by economists (see last row of Table 4.3).

This simple comparison of sample means for economists and non-economists, however,

Table 4.3: Proportional sacrifices for economists and non-economists

	Economists			Non-economists		
Treatment	$\overline{\rho}$	s.d	obs.	$\overline{\rho}$	s.d.	obs.
Non-med-FFS	0.2790	0.1135	10	0.2333	0.2746	12
Non-med-CAP	0.4470	0.2441	10	0.7316	0.2961	12
Overall	0.3630	0.2044	20	0.4824	0.3778	24

could be quite misleading. Thus, we investigate proportional sacrifices of Non-Meds grouped by economics and non-economics students for both payment conditions. Here, we find economists to be, on average, somewhat more caring for their patients, although not significant ($p = 0.1557$), than non-economists in treatment Non-Med-FFS (see Table 4.3). Further, Table 4.3 shows that economists' willingness to sacrifice own profit is lower than for non-economists under CAP. This difference is highly significant ($p = 0.0083$).

Thus, different proportional sacrifices between economists and non-economists under CAP indicate that behavioral differences between Meds and Non-Meds are driven by a more selfish behavior of economists among the Non-Meds. This finding is somewhat in line with Ahlert et al. (2008), although they employed a different experimental setup. Moreover, average proportional sacrifices of non-economists and Meds do not differ significantly under CAP ($p = 0.4362$, Fisher Pitman permutation test). In treatment Non-med-FFS we observed no behavioral differences between economists and non-economists. However, average sacrifices of Meds differ significantly from sacrifices of non-economists ($p = 0.0089$, Fisher Pitman permutation test).

Result 3. *Economists are significantly less willing to sacrifice own profit than non-economists in treatment Non-Med-CAP. This indicates that the more selfish behavior, observed under CAP, by Non-Meds compared to Meds is mainly driven by economists. However, proportional sacrifices between economists and non-economists do not differ significantly under FFS.*

4.3.3 Motivation

The education of medical students is supposed to be guided by the principles of respect for individuals, beneficence, and justice (Jagsi and Lehmann, 2004). Due to their medical education we hypothesized that Meds are more inclined to be motivated by the patient's welfare. The behavioral data of our experiment corroborated this conjecture. Moreover, we are able to check for further evidence on Meds patient-oriented motivations by the final questionnaire where we asked Meds and Non-Meds to state reasons and motivations for their choices in the following open question: "*Please recall the decision setting of the experiment. What influenced and motivated your decision? Why did you decide as you did?*".

Indeed, the patient benefit seems to decisively motivate Meds' decisions as it is mentioned by about 95% (40 out of 42) Meds. For the Non-Meds 77% (34 out of 44) mention the patient benefit. This difference is significant ($p = 0.0270$ Fisher exact test). As another determinant for their decisions 70 of the 86 subjects mention their own profit, i.e., about 71% of Meds and 92% Non-Meds. This implies that profit is mentioned significantly more often by Non-Meds than by meds ($p = 0.0270$ Fisher exact test). Further evidence for a more patient-oriented motivation of Meds is provided by only one medical student stating to be exclusively motivated by own profit. All other 41 Meds state to be motivated by the patient benefit only or both, profit and patient benefit. Of the Non-Meds, however, eight participants disclose to be only motivated by own profit.

In short, we find that Meds' decisions are substantially motivated by the patient's welfare. Non-Meds are, rather, motivated by their own profit.

4.4 Concluding remarks

The present study systematically analyzes behavior of prospective physicians (Meds) and non-medical students (Non-Meds) in a medical decision-making context. Our vehicle is a controlled laboratory experiment introduced by Hennig-Schmidt et al. (2009). Incentives from payment systems fee-for-service (FFS) and capitation (CAP) influence both, Meds and Non-Meds, in their decisions to provide medical services. In particular, we observe overprovision when subjects are paid by FFS and underprovision under CAP.

However, the extent to which subjects' quantity choices are influenced by payment incentives varies significantly across Meds and Non-Meds. Meds' choices are substantially influenced by the patient's benefit. This is reflected in a higher willingness to sacrifice own profit to increase the patient's benefit. Meds are less driven by profit maximizing objectives compared to Non-Meds. Put differently, Meds behave more altruisticaly than Non-Meds. In line with their behavior 95% of Meds state to be decisively influenced and motivated by the patients' benefit, in a questionnaire suceeding the experiment.

Behavioral data suggest that financial incentives work for Meds only in an alleviated way, as their behavior is mostly driven by patient-regarding (altrustic) concerns. Quantity choices maximizing the patient's welfare occur more frequently in CAP than in FFS for both Meds and Non-Meds. This indicates that more altruistic behavior can be observed under CAP.

More selfishness of economists in a medically framed experiment observed by Ahlert et al. (2008) can only partially be found in our behavioral data. In particular, economists are significantly less willing to sacrifice own profit than non-economists when paid by CAP and, thus, drive the behavioral differences across Meds and Non-Meds. Under FFS, however, proportional sacrifices between economists and non-economists do not differ significantly.

The findings from our study are not only relevant for the theoretical literature on optimal incentive structures in the public sector, but also suggest important policy implications for the health care sector. Incentives from payment systems negative for the patient's welfare are mitigated by the altruistic motivation of prospective physicians. Secondly, from a policy maker's point of view it is important to attract individuals with other-regarding motivations to work in the health care sector. If, however, the average participant from our Non-Med population would opt to work as a provider in this sector this would induce a substantial negative effect on patients' welfare. The possibility and significance of such effects must be considered carefully when designing robust incentive structures in the health care sector.

Bibliography

Ahlert, M., S. Felder, and B. Voigt (2008): "How Economists and Physicians Trade Off Efficiency and Equity in Medically and Neutrally Framed Allocation Problems?" FOR 655 Working Paper No. 09/2008, Jacobs University Bremen.

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström (2006): "Elicitation Using Multiple Price List Formats," *Experimental Economics*, 9, 383–405.

Arrow, K. (1963): "Uncertainty and the Welfare Economics of Medical Care," *American Economic Review*, 53, 941–969.

Arrow, K. J. (1965): *Aspects in the Theory of Risk-bearing*, Yrjö Jahnssonin Saeaetioe, Helsinki.

Ball, S. and P. Cech (1996): "Subject Pool Choice and Treatment Effects in Economic Laboratory Research," in *Research in Experimental Economics Vol. 6*, ed. by R. M. Isaac, JAI Press Inc., 239–292.

Baltussen, G., T. Post, M. J. van den Assem, and P. P. Wakker (2010): "Random Incentive Systems in a Dynamic Choice Experiment," Working Paper, University of Rotterdam.

Barr, A. M. and T. G. Packard (2002): "Revealed Preference and Self-Insurance: Can We Learn from the Self-Employed in Chile?" World Bank Policy Research Working Paper No. 2754.

Barros, P. (2003): "Cream-Skimming, Incentives for Efficiency and Payment System," *Journal of Health Economics*, 22, 419–443.

Barsky, R. B., M. S. Kimball, F. T. Juster, and M. D. Shapiro (1997): "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study," *Quarterly Journal of Economics*, 112, 537–579.

Beattie, J. and G. Loomes (1997): "The Impact of Incentives upon Risky Choice Experiments," *Journal of Risk and Uncertainty*, 14, 155–168.

Becker, G., M. DeGroote, and J. Marschak (1964): "Measuring Utility by a Single-Response Sequential Method," *Behavioral Science*, 9, 226–232.

Besley, T. and M. Ghatak (2005): "Competition and Incentives with Motivated Agents," *American Economic Review*, 95, 616–636.

Binswanger, H. P. (1980): "Attitudes Towards Risk: Experimental Measurement in Rural India," *American Journal of Agricultural Economics*, 62, 395–407.

Blomqvist, Å. (1991): "The Doctor as Double Agent: Information Asymmetry, Health Insurance, and Medical Care," *Journal of Health Economics*, 10, 411–432.

Briec, W., K. Kerstens, and O. Jokung (2007): "Mean-Variance-Skewness Portfolio Performance Gauging: A General Shortage Function and Dual Approach," *Management Science*, 53, 135–149.

Brockett, P. L. and J. R. Garven (1998): "A Reaxamination of the Relationship Between Preferences and Moment Orderings by Rational Risk-Averse Investors," *Geneva Papers Risk Insurance Theory*, 23, 127–137.

Brockett, P. L. and L. L. Golden (1987): "A Class of Utility Functions Containing All the Common Utility Functions," *Management Science*, 33, 955–964.

Brockett, P. L. and Y. Kahane (1992): "Risk, Return, Skewness and Preference," *Management Science*, 38.

Bruner, D. (2009): "Changing the Probability versus Changing the Reward," *Experimental Economics*, 12, 367–385.

Caballe, J. and A. Pomansky (1996): "Mixed Risk Aversion," *Journal of Economic Theory*, 71, 485–513.

Cadsby, C. and E. Maynes (1998): "Choosing between a Socially Efficient and Free-riding Equilibrium: Nurses versus Economics and Business Students," *Journal of Economic Behavior and Organization*, 37, 183–192.

Camerer, C. F. (1989): "An Experimental Test of Several Generalized Utility Theories," *Journal of Risk and Uncertainty*, 2, 61–104.

Camerer, C. F. and R. M. Hogarth (1999): “The Effects of Financial Incentives in Economics Experiments: A Review and Capital-production-labor Framework,” *Journal of Risk and Uncertainty*, 18, 7–42.

Carrol, C. D. (1994): “How does Future Income Affect Current Consumption?” *Quarterly Journal of Economics*, 109, 111–147.

Carroll, C. D. and M. S. Kimball (2008): “Precautionary Savings and Precautionary Wealth,” in *The New Palgrave Dictionary of Economics*, ed. by S. Durlauf and L. Blume, MacMillan, London.

Carter, J. and M. Irons (1991): “Are Economists Different, and if so, why?” *Journal of Economic Perspectives*, 5, 171–177.

Chiu, H. (2005): “Skewness Preference, Risk Aversion and the Precedence Relations on Stochastic Changes,” *Management Science*, 51, 1816–28.

——— (forthcoming): “Skew Preference, Risk Taking and Expected Utility Maximization,” *Geneva Risk and Insurance Review*.

Choné, P. and C. A. Ma (2010): “Optimal Health Care Contract under Physician Agency,” Working Paper, University of Boston.

Courbagé, C. and B. Rey (2006): “Prudence and Optimal Prevention for Health Risks,” *Health Economics*, 15, 1323–1327.

Crainich, D. and L. Eeckhoudt (2008): “On the Intensity of Downside Risk Aversion,” *Journal of Risk and Uncertainty*, 36, 267–276.

Croson, R. T. A. and U. Gneezy (2009): “Gender Differences in Preferences,” *Journal of Economic Literature*, 47, 448–474.

Cubitt, R., C. Starmer, and R. Sugden (1998): “On the Validity of the Random Lottery Incentive System,” *Experimental Economics*, 1, 115–131.

Davidson, S., L. Manheim, S. Werner, M. Hohlen, B. Yudkowsky, and G. Flemming (1992): “Prepayment with Office-based Physicians in Publicly Funded Programs: Results from Children’s Medicaid Program,” *Pediatrics*, 89, 761–767.

Davis, D. D. and C. A. Holt (1993): *Experimental Economics*, Princeton University Press, Princeton (N.J.).

De Jaegher, K. and M. Jegers (2000): "A Model of Physician Behaviour with Demand Inducement," *Journal of Health Economics*, 19, 231–258.

Deck, C. A. and H. Schlesinger (2010): "Exploring Higher Order Risk Effects," *Review of Economic Studies*, 77, 1403–1420.

Delfgaauw, J. (2007): "Dedicated Doctors: Public and Private Provision of Health Care with Altruistic Physicians," University of Rotterdam, Tinbergen Institute Discussion Papers, No. 2007-010/1.

Delfgaauw, J. and R. Dur (2008): "Incentives and Workers' Motivations in the Public Sector," *Economic Journal*, 118, 171–191.

Denuit, M. M. and L. Eeckhoudt (forthcoming a): "Stronger Measures of Higher-order Risk Attitudes," *Journal of Economic Theory*.

——— (forthcoming b): "A General Index of Absolute Risk Attitude," *Management Science*.

Devlina, R. and S. Sarma (2008): "Do Physician Remuneration Schemes Matter? The Case of Canadian Family Physicians," *Journal of Health Economics*, 27, 1168–1181.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde (forthcoming a): "Are Risk Aversion and Impatience Related to Cognitive Ability?" *American Economic Review*.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (forthcoming b): "Individual Risk Attitudes: Measurement, Determinants and Behavioral Consequences," *Journal of the European Economic Association*.

Dulleck, U. and R. Kerschbamer (2006): "On Doctors, Mechanics and Computer Specialists – The Economics of Credence Goods," *Journal of Economic Literature*, 44, 5–42.

Dulleck, U., R. Kerschbamer, and M. Sutter (forthcoming): "The Economics of Credence Goods: An Experiment on the Role of Liability, Verifiability, Reputation and Competition," *American Economic Review*.

Dumont, E., B. Fortin, N. Jacquement, and B. Shearer (2008): "Physicians' Multitasking and Incentives: Empirical Evidence from a Natural Experiment," *Journal of Health Economics*, 27, 1436–1450.

Dynan, K. E. (1993): "How Prudent are Consumers," *Journal of Political Economy*, 101, 1104–1013.

Ebert, S. (2010): "Moment Characterization of Higher-order Risk Preferences," University of Bonn, Bonn Econ Discussion Papers, No. 17/2010.

Ebert, S. and D. Wiesen (2009): "An Experimental Methodology Testing for Prudence and Third-order Preferences," University of Bonn, Bonn Econ Discussion Papers, No. 21/2009.

Eckel, C. C. and P. J. Grossman (1996): "Altruism in Anonymous Dictator Games," *Games and Economic Behavior*, 16, 181–191.

——— (2008 a): "Differences in the Economic Decisions of Men and Women: Experimental Evidence," in *Handbook of Experimental Economics Results Vol. 1*, ed. by C. Plott and V. Smith, Elsevier, New York, 1061–1073.

——— (2008 b): "Forecasting Risk Attitudes: An Experimental Study Using Actual and Forecast Gamble Choices," *Journal of Economic Behavior and Organization*, 68, 1–17.

Eeckhoudt, L. and C. Gollier (2005): "The Impact of Prudence on Optimal Prevention," *Economic Theory*, 26, 989–994.

Eeckhoudt, L., B. Rey, and H. Schlesinger (2007): "A Good Sign for Multivariate Risk Taking," *Management Science*, 53, 117–124.

Eeckhoudt, L. and H. Schlesinger (2006): "Putting Risk in Its Proper Place," *American Economic Review*, 96, 280–289.

——— (2008): "Changes in Risk and the Demand for Saving," *Journal of Monetary Economics*, 55, 1329–1336.

Eichner, T. (2008): "Mean Variance Vulnerability," *Management Science*, 54, 586–593.

Ekern, S. (1980): "Increasing Nth Degree Risk," *Economics Letters*, 6, 329–333.

Ellis, R. and T. McGuire (1986): "Provider Behavior under Prospective Reimbursement: Cost Sharing and Supply," *Journal of Health Economics*, 5, 129–151.

Esö, P. and L. White (2004): "Precautionary Bidding in Auctions," *Econometrica*, 72, 77–92.

Fagart, M. and B. Sinclair-Desgagné (2007): “Ranking Contingent Monitoring Systems,” *Management Science*, 53, 1501–1509.

Falk, A. and E. Fehr (2003): “Why Labor Market Experiments?” *Labor Economics*, 10, 399–406.

Fan, C., K. Chen, and K. Kan (1998): “The Design of Payment Systems for Physicians under Global Budget - An Experimental Study,” *Journal of Economic Behavior and Organization*, 34, 295–311.

Fei, W. and H. Schlesinger (2008): “Precautionary Insurance Demand with State-dependent Background Risk,” *Journal of Risk and Insurance*, 75, 1–16.

Fischbacher, U. (2007): “Z-tree: Zurich Toolboox for Readymade Economic Experiments,” *Experimental Economics*, 10, 171–178.

Fischbacher, U., S. Gächter, and E. Fehr (2001): “Are People Conditionally Cooperative? Evidence from a Public Goods Experiment,” *Economics Letters*, 71, 397–404.

Francois, P. (2000): “Public Service Motivation as an Argument for Government Provision,” *Journal of Public Economics*, 78, 275–299.

Frank, R. (2007): “Behavioral Economics and Health Economics,” in *Behavioral Economics and Its Applications*, ed. by P. Diamond and H. Vartianinen, Princeton University Press, New Jersey, 195–234.

Fuchs, V. R. (2000): “The Future of Health Economics,” *Journal of Health Economics*, 19, 141–157.

Gilboa, I. (2009): *Theory of Decision under Uncertainty*, Econometric Society Monographs ESM 45, Cambridge University Press, New York.

Gollier, C. (2001): *The Economics of Risk and Time*, MIT Press, Cambridge (MA).

——— (2010): “Ecological discounting,” *Journal of Economic Theory*, 145, 812–829.

Gomes, F. and A. Michaelides (2005): “Optimal Life-Cycle Asset Allocation: Understanding the Empirical Evidence,” *Journal of Finance*, 60, 869–904.

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

Gosden, T., F. Forland, I. Kristiansen, M. Sutton, B. Leese, A. Guiffrida, M. Sergison, and L. Pedersen (2001): "Impact of Payment Method on Behavior of Primary Care Physicians: A Systematic Review," *Journal of Health Services Research and Policy*, 6, 44–54.

Greiner, B. (2004): "An Online Recruitment System for Economic Experiments," in *Forschung und wissenschaftliches Rechnen : Beiträge zum Heinz-Billing-Preis 2003*, ed. by K. Kremer and V. Macho, Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen, 79–93.

Grether, D. M. and C. R. Plott (1979): "Economic Theory of Choice and the Preference Reversal Phenomenon," *American Economic Review*, 69, 623–638.

Grytten, J. and R. Sørensen (2001): "Type of Contract and Supplier-induced Demand for Primary Physicians in Norway," *Journal of Health Economics*, 20, 379–393.

Harrison, G. W. (1986): "An Experimental Test for Risk Aversion," *Economics Letters*, 21, 7–11.

Harrison, G. W., E. Johnson, M. M. McInnes, and E. E. Rutström (2005): "Risk Aversion and Incentive Effects: Comment," *American Economic Review*, 95, 897–901.

Harrison, G. W., J. A. List, and C. Towe (2007): "Naturally Occurring Preferences and Exogenous Laboratory Experiments: A Case Study of Risk Aversion," *Econometrica*, 75, 433–458.

Harrison, G. W. and E. E. Rutström (2008): "Risk Aversion in the Laboratory," in *Research in Experimental Economics*, ed. by G. Harrison and J. Cox, JAI Press, Greenwich (CT).

Hennig-Schmidt, H., R. Selten, and D. Wiesen (2009): "How Payment Systems Affect Physicians' Provision Behavior – An Experimental Investigation," University of Bonn, Bonn Econ Discussion Paper Series, No. 29/2009.

Hennig-Schmidt, H. and D. Wiesen (2010): "Are Prospective Doctors Different? Evidence from an Experimental Study on a Payment System Variation," Working paper, University of Bonn.

Hertwig, R. and A. Ortmann (2001): "Experimental Practices in Economics: A

Methodological Challenge for Psychologists?" *Behavioral and Brain Sciences*, 24, 383–403.

Hey, J. D. and C. D. Orme (1994): "Investigating Generalizations of Expected Utility Theory Using Experimental Data," *Econometrica*, 62, 1291–1326.

Holt, C. A. and S. K. Laury (2002): "Risk Aversion and Incentive Effects," *American Economic Review*, 92, 1644–1655.

Hutchinson, B., S. Birch, J. Hurley, J. Lomas, and F. Stratford-Devai (1996): "Do Physician-payment Mechanisms Affect Hospital Utilization? A Study of Health Service Organizations in Ontario," *Canadian Medical Association Journal*, 154, 653–661.

Iversen, T. (1993): "A Theory of Hospital Waiting Lists," *Journal of Health Economics*, 12, 55–71.

Iversen, T. and H. Lurås (2000): "The Effect of Capitation on GPs' Referal Decision," *Health Economics*, 9, 199–210.

Jagsi, R. and L. Lehmann (2004): "The Ethics of Medical Education," *British Medical Journal*, 329, 332–334.

Jindapon, P. and W. Neilson (2007): "Higher-order Generalixations of Arrow-Pratt and Ross Risk Aversion: A Comparative Statics Approach," *Journal of Economic Theory*, 136, 719–728.

Kahneman, D., J. L. Knetsch, and R. Thaler (1986): "Fairness and Assumptions of Economics," *Journal of Business*, 59, 285–300.

Kahneman, D. and A. Tversky (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47, 263–292.

Kimball, M. S. (1990): "Precautionary Savings in the Small and in the Large," *Econometrica*, 58, 53–73.

——— (1992): "Precautionary Motives for Holding Assets," in *The New Palgrave Dictionary of Money and Finance*, ed. by P. Newman, M. Milgate, and e. John Eatwell, Stockton Press, New York, 158–161.

——— (1993): "Standard Risk Aversion," *Econometrica*, 61, 589–611.

Krasnik, A., P. Groenewegen, P. Pedersen, P. von Scholten, G. Mooney,

A. Gottschau, H. Flierman, and M. Damsgaard (1990): "Changing Remuneration Systems: Effects on Activity in General Practice," *British Medical Journal*, 300, 1698–1701.

Kraus, A. and R. Litzenberger (1976): "Skewness Preference and the Valuation of Risky Assets," *Journal of Finance*, 31.

Kuon, B. (1994): *Two-Person Bargaining Experiments with Incomplete Information*, Lecture Notes in Economics and Mathematical Systems, Vol. 412, Springer, New York.

Laury, S. K. (2005): "Pay One or Pay All: Random Selection of One Choice for Payment," Working Paper, Andrew Young School of Policy Studies Research Paper Series 06-13.

Le Grand, J. (2003): *Motivation, Agency, and Public Policy : Of Knights and Knaves, Pawns and Queens*, Oxford University Press, Oxford (UK).

Leland, H. E. (1968): "Saving and Uncertainty: The Precautionary Demand for Saving," *Quarterly Journal of Ecoomics*, 82, 465–473.

Lévy-Garboua, L., C. Montmarquette, and M. Villeval (2008): "Individual Responsibility and Public Funding of Health Care: An Experiment," Working paper, CIRANO University of Montreal.

Loomes, G. (1988): "Different Experimental Procedures for Obtaining Valuations of Risky Actions: Implications for Utility Theory," *Theory and Decision*, 25, 1–23.

Ma, C. A. (1994): "Health Care Payment Systems: Cost and Quality Incentives," *Journal of Economics and Management Strategy*, 3, 93–112.

——— (2007): "Altruism and Incentives in Public and Private Health Care," in *Finance and Incentives of the Health Care System, Proceedings of the 50th Anniversary Symposium of the Yrjö Jahnsson Foundation*, ed. by A. Suvanto and H. Vartiainen, Government Institute for Economic Research, Helsinki, 79–104.

MacGillivray, H. L. (1986): "Skewness and Asymmetry: Measures and Orderings," *Annals of Statistics*, 14, 994–1011.

Mao, J. (1970): "Survey of Capital Budgeting: Theory and Practice," *Journal of Finance*, 25, 349–369.

Marwell, G. and R. Ames (1981): "Economists Free Ride, Does Anyone Else?" *Journal of Public Economics*, 15, 259–310.

McGuire, T. G. (2000): "Physician Agency," in *Handbook of Health Economics, Vol. 1 A*, ed. by Cuyler and Newhouse, North-Holland, Amsterdam, 461–536.

Menezes, C. F., C. Geiss, and J. Tressler (1980): "Increasing Downside Risk," *American Economic Review*, 70, 921–932.

Menezes, C. F. and X. H. Wang (2005): "Increasing Outer Risk," *Journal of Mathematical Economics*, 41, 875–886.

Modica, S. and M. Scarsini (2005): "A Note on Comparative Downside Risk Aversion," *Journal of Economic Theory*, 122, 267–271.

Montgomery, D. C. (2005): *Design and Analysis of Experiments*, John Wiley & Sons, New York.

Murnighan, J. K., A. Roth, and F. Schoumaker (1988): "Risk Aversion in Bargaining: An Experimental Study," *Journal of Risk and Uncertainty*, 1, 101–24.

Newhouse, J. P. (1996): "Reimbursing Health Plans and Health Providers: Efficiency in Production Versus Selection," *Journal of Economic Literature*, 34, 1236–1263.

——— (2002): *Pricing the Priceless – A Health Care Conundrum*, MIT Press, Cambridge (MA).

Newhouse, J. P. and the Insurance Experiment Group (1993): *Free for All: Lessons from the RAND Health Insurance Experiment*, Harvard University Press, Cambridge (MA).

Olivella, P. and M. Vera-Hernandez (2007): "Competition among Differentiated Health Plans under Adverse Selection," *Journal of Health Economics*, 26, 233–250.

Plott, C. R. and V. L. Smith, eds. (2008): *Handbook of Experimental Economics Results*, vol. 1, Elsevier, New York.

Pratt, J. W. (1964): "Risk Aversion in the Small and in the Large," *Econometrica*, 32, 122–136.

Pratt, J. W. and R. J. Zeckhauser (1987): "Proper Risk Aversion," *Econometrica*, 55, 143–154.

Roger, P. (forthcoming): "Mixed Risk Aversion and Preference for Risk Disaggregation: A Story of Moments," *Theory and Decision*.

Ross, S. A. (1981): "Some Stronger Measures of Risk Aversion in the Small and in the Large with Applications," *Econometrica*, 49, 621–663.

Rothschild, M. D. and J. E. Stiglitz (1970): "Increasing Risk: I. A Definition," *Journal of Economic Theory*, 2(3), 225–243.

Sandmo, A. (1970): "The Effect of Uncertainty on Saving Decisions," *Review of Economic Studies*, 37, 353–360.

Schram, A. and J. Sonnemans (2008): "How Individuals Choose Health Insurance: An Experimental Analysis," Working paper, University of Amsterdam.

Schubert, R., M. Brown, M. Gysler, and H. W. Brachinger (1999): "Financial Decision-Making: Are Women Really More Risk-Averse?" *American Economic Review*, 89, 381–385.

Selten, R. (1967): "Ein Oligopolexperiment mit Preisvariation und Investition," in *Beiträge zur experimentellen Wirtschaftsforschung*, ed. by H. Sauermann, J.C.B. Mohr (Paul Siebeck), Tübingen, 103–135.

Selten, R. and A. Ockenfels (1998): "An Experimental Solidarity Game," *Journal of Economic Behavior and Organization*, 34, 517–539.

Starmer, C. and R. Sugden (1991): "Does the Random-Lottery Incentive System Elicit True Preferences? An Experimental Investigation," *American Economic Review*, 81, 971–78.

Stearns, S., B. Wolfe, and D. Kindig (1992): "Physician Responses to Fee-for-service and Capitation Payment," *Inquiry*, 29, 416–425.

Tarazona-Gomez, M. (2004): "Are Individuals Prudent? An Experimental Approach Using Lottery Choices," Working paper, Copenhagen Business School.

Treich, N. (forthcoming): "Risk-aversion and Prudence in Rent-seeking Games," *Public Choice*.

Tsetlin, I. and R. L. Winkler (2009): "Multiattribute Utility Satisfying a Preference for Combining Good with Bad," *Management Science*, 55, 1942–1952.

van Zwet, W. (1964): "Convex Transformations of Random Variables," Mathematical Centre Tracts 7, Mathematisch Centrum, Amsterdam.

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

White, L. (2008): "Prudence in Bargaining: The effect of uncertainty on bargaining outcomes," *Games and Economic Behavior*, 62, 211–231.

Woodward, R. S. and F. Warren-Boulton (1984): "Considering the Effect of Financial Incentives and Professional Ethics on 'Appropriate' Medical Care," *Journal of Health Economics*, 3, 223–237.

Appendices

A.1 Appendix to Chapter 1

A.1.1 Proofs

Proof of Proposition 2. Let $n \in \mathbb{N}$ be arbitrary. It is well known that the $\mu_n(\cdot)$-operator is homogeneous of degree n and translation invariant. The assumption $p_X = 1 - p_Y$ is equivalent to $X = 1 - Y$ such that $\mu_n(X) = (-1)^n \mu_n(Y)$ which for $n = 2$ just implies $\mathbb{V}(X) = \mathbb{V}(Y)$. Note that we can write $L_X = X \cdot x_1 + (1 - X) \cdot x_0 = (x_1 - x_0)X + x_0$ and thus $\mathbb{V}(L_X) = (x_1 - x_0)^2 \mathbb{V}(X)$. Analogously, we have $\mathbb{V}(L_Y) = (y_1 - y_0)^2 \mathbb{V}(Y)$. Since the Mao lotteries have equal variance we obtain $(x_1 - x_0)^2 = (y_1 - y_0)^2$ and because of the unique representation of binary lotteries (see Definition 1) this is equivalent to

$$x_1 - x_0 = y_1 - y_0. \tag{1}$$

Using once more homogeneity and translation invariance of the $\mu_n(\cdot)$-operator and plugging in yields

$$\mu_n(L_X) = (x_1 - x_0)^n \mu_n(X) = ((y_1 - y_0))^n (-1)^n \mu_n(Y) = (-1)^n \mu_n(L_Y).$$

Because of the assumed variance equality the claim for $\mu_n^S(\cdot)$ follows immediately. □

Proof of Proposition 3. The proof can be done by statistical standard calculations and thus is omitted. A more general version of the proposition is proven as Proposition 1 in Ebert (2010). □

Proof of Theorem 1. After calculating expectation, variance and skewness of a binary lottery as in Definition 1 we find that $L_X = L_X(p, x_1, x_0)$ has to suffice the following system of equations

$$E = px_1 + (1-p)x_0 \quad (2)$$
$$V = (x_1 - x_0)^2 p(1-p) \quad (3)$$
$$S = \frac{1-2p}{\sqrt{p(1-p)}} \quad (4)$$

with $x_1 > x_0$ and $0 < p < 1$. It is natural to start with solving equation (4) for p. After squaring and some rearrangement one obtains

$$p^2(-S^2 - 4) + p(4 + S^2) - 1 = 0.$$

Setting $\tilde{S} := 4 + S^2$ the solutions to this quadratic equation are given by

$$p_{1/2} = \frac{\tilde{S} \pm \sqrt{\tilde{S}^2 - 4\tilde{S}}}{2\tilde{S}}, \quad (5)$$

where p_1 is the solution associated with the addition. It is easy to see that the expression under the square root is always positive. If $S = 0$ there is one solution, namely $p = \frac{1}{2}$. Otherwise there are two solutions. Both these solutions are strictly positive since $\sqrt{\tilde{S}^2 - 4\tilde{S} + 4 - 4} = \sqrt{(\tilde{S} - 2)^2 - 4)} \leq \tilde{S} - 2$ and thus

$$p_{1/2} \geq p_2 \geq \frac{\tilde{S} - (\tilde{S} - 2)}{2\tilde{S}} = \frac{1}{\tilde{S}} > 0.$$

All solutions are smaller than 1 since

$$p_1 < 1 \Longleftrightarrow \sqrt{\tilde{S}^2 - 4\tilde{S}} < \tilde{S}$$

what can be shown to be true for all $\tilde{S}$ (and thus for all S) by doing the quadratic expansion as in the previous calculation. Note that equation (4) is a square root equation and thus we have to verify the solutions. Obviously, if $S = 0$ then $p = 0.5$ is the unique solution. Otherwise, from equation (5) it follows that $p_1 > p_2$ and $p_1 + p_2 = 1$, i.e., $p_1 > 0.5$ and $p_2 < 0.5$. If $S < 0$ then p_2 does not solve equation (4) because $1 - 2p_2 > 0$, but p_1 does. Similarly, if $S > 0$ only p_2 solves equation (4). Thus in any case equation (4) has a unique solution in $(0, 1)$ (such that it is a probability) that will be denoted by p in the following.[7]

[7] We can see now how skewness is reflected in a binary lottery. It has zero skewness if and only if both states have equal probability. Otherwise, it is positively (negatively) skewed if the high payoff is associated with the low (high) probability.

The remainder of the proof is straightforward. For any p obtained from equation (4) the system of equations (2) and (3) can be solved for a unique solution to obtain the expressions stated in the claim from which finally also $x_1 > x_0$ is evident. □

Proof of Proposition 4. By Theorem 1 there exists exactly one binary lottery $L_X \equiv M_A$ with $\mathbb{E}[M_A] = \mathbb{E}[A]$, $\mathbb{V}[M_A] = \mathbb{V}[A]$ and $\nu[M_A] = -S$. By Theorem 1 there also exists exactly one lottery $L_Y \equiv M_B$ whose expectation and variance equal that of M_A and further $\nu[M_B] = +S$. From equation (4) one immediately obtains $p_X = 1 - p_Y$ such that by Definition 2 (M_A, M_B) constitutes a Mao lottery pair that fulfills the requested moment conditions.

For the second part, note that by taking derivatives

$$\Delta = (\nu[B] - S)^2 + (\nu[A] - (-S))^2$$

indeed achieves its maximum at $S = \frac{\nu[B]-\nu[A]}{2}$. The difference in skewness of the Mao pair is $2S$ and as can be seen from the previous equation this indeed equals the skewness difference $\nu[B] - \nu[A]$ of the prudence pair. □

A.1.2 Instructions

[*translated from German for session order MAO-ES-RIAV*]

Thank you very much for participating in this decision experiment!

General Information

In the following experiment, you will make a couple of decisions. Following the instructions and depending on your decisions, you can earn money. It is therefore very important that you read the instructions carefully.

You will make your decisions anonymously on your computer screen in your cubicle. During the experiment you are not allowed to talk to the other participants. Whenever you have a question, please raise your hand. The experimenter will answer your question in private in your cubicle. If you disregard these rules you can be excluded from the experiment. Then you receive no payment.

During the experiment all amounts are stated in Taler, the experimental currency. At the end of the experiment, your achieved earnings will be converted into Euro at an exchange rate of 1 Taler = € 0.15 and paid to you in cash.

Structure of the Experiment and Your Decisions

In total, you will make 34 decisions throughout the experiment. In each decision you will decide upon which of **two different risky events—either Option A or Option B**—you prefer. An example of Option A could be as follows: *With 50% chance you will lose 10 Taler or with 50% chance you will receive 20 Taler. Option B could be: With 20% chance you will receive 0 Taler and with 80% chance you will receive 10 Taler.*

The experiment consists of three stages that will be explained in detail in the following. To determine your payoff in the experiment, one of your decisions will be randomly chosen. This takes place after you have completed all your decisions. To this end, the experimenter picks one of 34 balls, marked with numbers from 1 to 34, out of a ballot box. Each number occurs only once in the ballot box, whereby the draw of a particular number is equally likely. The outcome of the risky event, that you have opted for, at the randomly chosen decision will afterwards be determined by another random draw. This procedure will be explained extensively when the stages of the experiment are described.

Keep in mind that only one of your 34 decisions determines your payoff in the experiment. Therefore each of your single decisions can determine your entire

payoff in the experiment.

You make your decisions at the computer screen in the computer lab. For each decision you have a maximum of 3 minutes. After the experiment, the decision relevant for every participants's payoff and the outcome of the risky event will be determined by random draws for each participant in the seminar room. For this the experimenter will call upon participants one by one.

Note that some risky events comprise negative outcomes. For these questions you receive coupons indicating an endowment (in Taler). You can charge the coupons when the outcome of the risky event is determined.

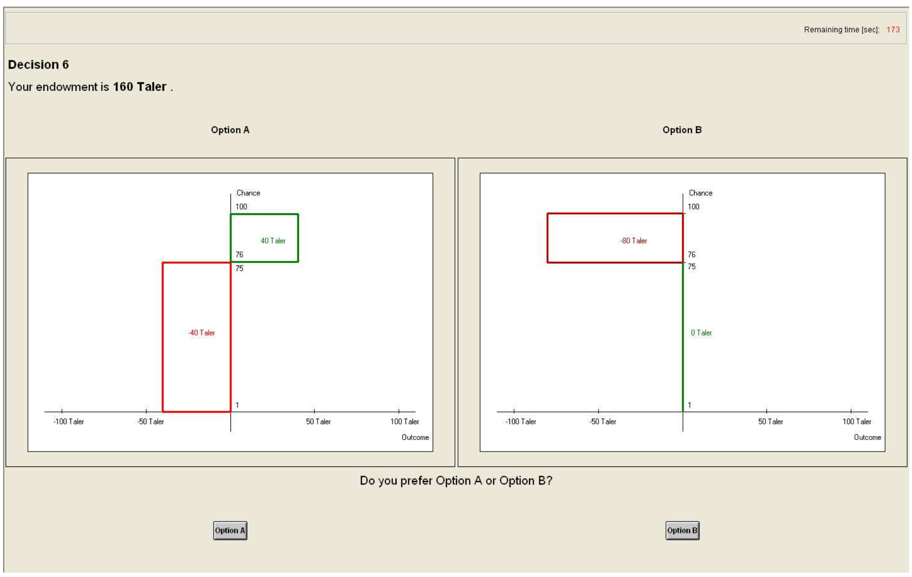
Stage I

In the first stage of the experiment you are asked to make eight decisions. On each of the 8 sequent decision screens you decide which of the two risky events—Option A or Option B—you prefer.

For your decisions you receive an endowment in Taler, because outcomes of risky events in this stage can comprise losses. Accordingly, your payoff in this stage is:

Endowment and **Outcome of the chosen risky event**

How is the outcome of the (chosen) risky event determined in Stage I? To this end, there is another ballot box. This ballot box contains 100 balls with numbers from 1 to 100. Each number occurs only once, thus the draw of a particular number is equally likely. An example of a decision screen provides the following screen:



In Option A you will lose 40 Taler with 75% chance (balls 1 to 75) or with 25% chance you will receive 40 Taler (balls 76 to 100). In Option B you receive 0 Taler with 75% chance (balls 1 to 75) or you will lose 80 Taler with 25% chance (balls 76 to 100). Your endowment is 160 Taler in this example.

Now suppose that this decision was randomly drawn to determine your payoff.

- *Suppose you have chosen* ***Option A*** *and assume that a ball is drawn from the ballot box with a number between 1 and 75. That means, you lose 40 Taler. Your resulting payoff, after allocating the endowment of 160 Taler for this decision to the lottery outcome, is 120 Taler. If a ball with a number between 76 and 100 is drawn, you receive 40 Taler. Under consideration of your endowment your payoff is 200 Taler.*
- *Suppose you have chosen* ***Option B*** *and assume that a ball is drawn from the ballot box with a number between 1 and 75. That means, you receive 0 Taler. Your resulting payoff after allocating the endowment of 160 Taler for this decision to the lottery outcome is 160 Taler. If a ball with a number between 76 and 100 is drawn, you lose 80 Taler. Under consideration of your endowment your payoff is 80 Taler.*

Stage II

In the second stage of the experiment you make 16 decisions. Again, on each of the 16 sequent decision screens you decide which of the two risky events—either Option A or Option B—you prefer. In this stage risky events (may) comprise two random draws.

For each decision one random draw is given. This draw is as follows: With 50% chance the situation "Up" occurs or with 50% chance the situation "Down" occurs.

For your decisions you receive an endowment in Taler, because outcomes of risky events in this stage can also comprise losses. Accordingly, your payoff in this stage consists of two components:

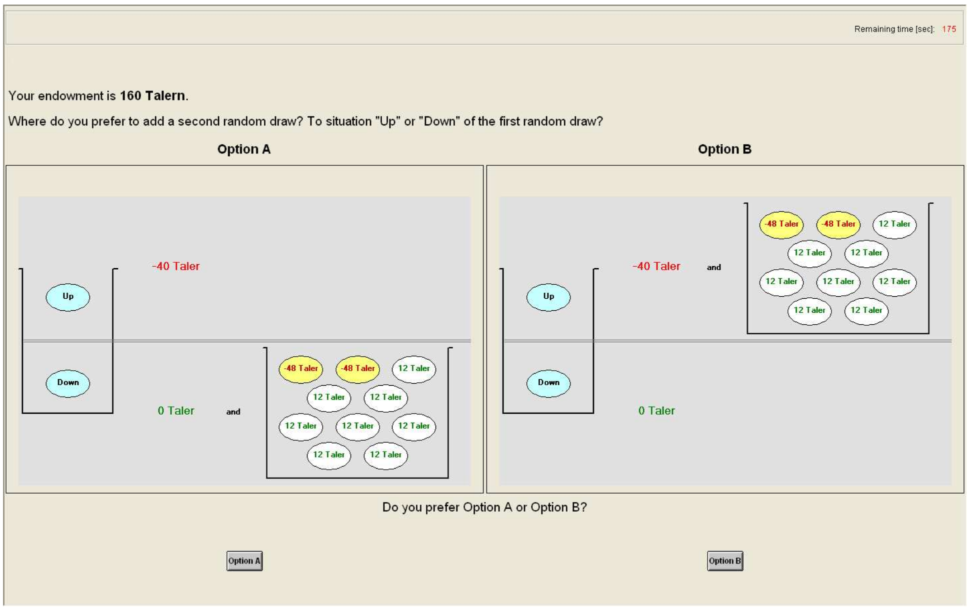
Endowment and **Outcome of the chosen risky event**

How is the outcome of the (chosen) risky event determined in Stage II? For the first random draw there are two balls in a ballot box—one marked with "Up" and the other with "Down". The draw of a particular ball is equally likely.
To determine your payoff in this stage **two** random draws may be necessary. At the second random draw one ball is drawn from another ballot box with 10 balls. The balls are either yellow or white. Note that the composition of yellow and white balls may change for different decisions in this stage. But within one decision, i.e., for Option A and Option B,

the composition of yellow and white balls remains the same.

Decision type 1

For 8 out of 16 decisions you are asked the following: Given what situation of the first random draw—either "Up" or "Down"—do you prefer a **second random draw**? An example is provided by the following screen:



In Option A you lose 40 Taler, if situation "Up" occurs in the first random draw. If situation "Down" occurs, you receive 0 Taler and a second random draw succeeds. This second random draw is as follows: With 20% chance you lose 48 Taler and with 80% chance you receive 12 Taler. In Option B, you lose 40 Taler if in the first random draw the situation "Up" occurs and a second random draw succeeds (The second random draw is the same as in Option A). When situation "Down" occurs, you receive 0 Taler. For this decision you are endowed with 160 Taler.

Now suppose the decision from the example above is randomly drawn to determine your payoff.
Suppose you have chosen ***Option A****.*

- *If in the first random draw the ball "Up" is drawn, you lose 40 Taler. After allocating your endowment of 160 Taler for this decision to the lottery outcome, your payoff is 120 Taler.*
- *If in the first random draw the ball "Down" is drawn, you receive 0 Taler and a second random draw succeeds.*
  - *If in the second random draw a yellow ball is drawn, you lose 48 Taler and your payoff after allocating your endowment is 112 Taler.*

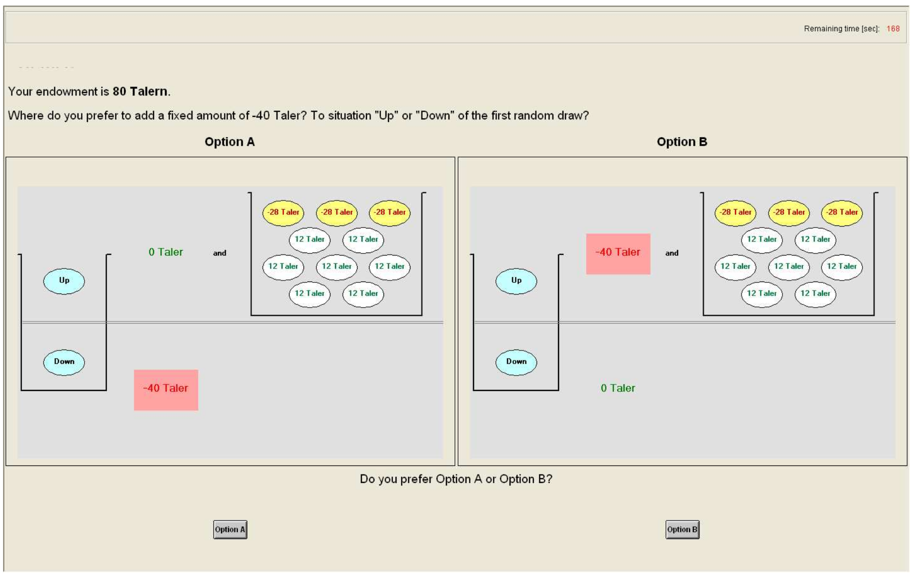
– *If in the second random draw a <u>white</u> ball is drawn, you receive 12 Taler and your payoff after allocating your endowment is 172 Taler.*

Suppose you have chosen ***Option B****.*

- *If in the first random draw ball "Up" is drawn, you lose 40 Taler and a second random draw succeeds.*
  – *If in the second random draw a <u>yellow</u> ball is drawn, you lose 48 Taler and your payoff after allocating your endowment is 72 Taler.*
  – *If in the second random draw a <u>white</u> ball is drawn, you receive 12 Taler and your payoff after allocating your endowment is 132 Taler.*
- *If in the first random draw "Down" is drawn, you receive 0 Taler. After allocating your endowment of 160 Taler for this decision to the lottery outcome your payoff is 160 Taler.*

Decision type 2

For the remaining 8 out of 16 decisions in stage II you are asked the following: To what situation do you prefer to add a **(fixed) amount**—either to situation "Up" where a second random draw succeeds or to situation "Down", where no second random draw succeeds. Note that the fixed amount can either be positive or negative. An example is provided by the following screen:



In Option A, when situation "Up" occurs in the first random draw you receive 0 Taler and a second random draw succeeds. The second random draw is as follows: With 30% chance you lose 28 Taler and with 70% chance you receive 12 Taler. When situation "Down" occurs *in the first random draw, you lose 40 Taler and no second random draw succeeds. In Option B if situation "Up" occurs in the first random draw you lose 40 Taler and a second random draw succeeds (The second random draw is the same as in Option A). When situation "Down" occurs, you receive 0 Taler and no second random draw succeeds. For this decision you are endowed with 80 Taler.*

*Now suppose the decision from the example above is randomly drawn to determine your payoff. Suppose you have chosen **Option A**.*

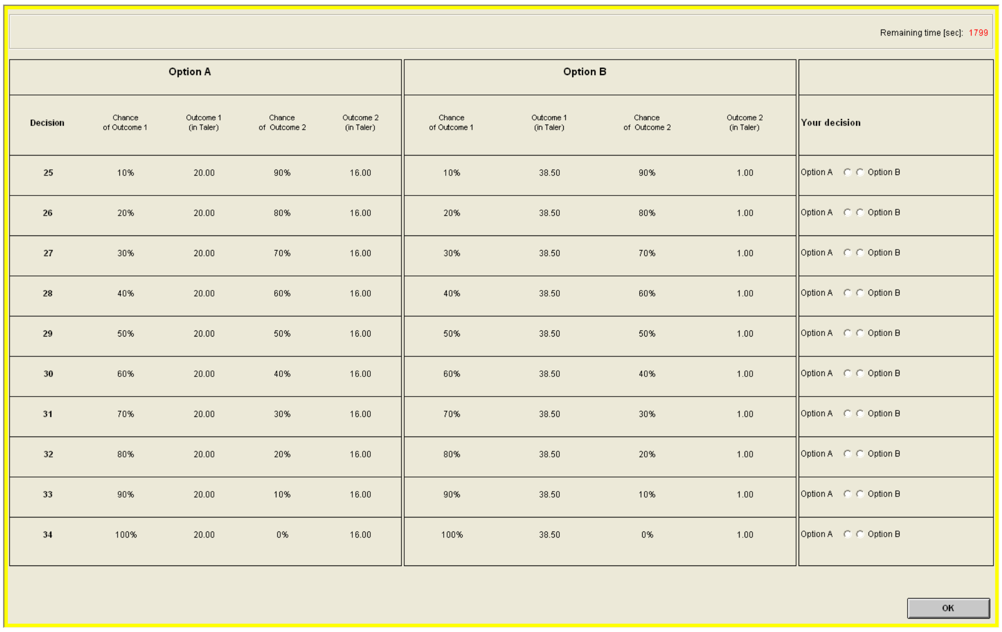
- *If in the first random draw the ball "Up" is drawn, you receive 0 Taler and a second random draw succeeds.*
  - *If in the second random draw a yellow ball is drawn, you lose 28 Taler and your payoff after allocating your endowment is 52 Taler.*
  - *If in the second random draw a white ball is drawn, you receive 12 Taler and your payoff after allocating your endowment is 92 Taler.*
- *If in the first random draw the ball "Down" is drawn, you lose 40 Taler. After allocating your endowment of 80 Taler for this decision to the lottery outcome your payoff is 40 Taler.*

*Suppose you have chosen **Option B**.*

- *If in the first random draw the ball "Up" is drawn, you lose 40 Taler and a second random draw succeeds.*
  - *If in the second random draw a yellow ball is drawn, you lose 28 Taler and your payoff after allocating your endowment is 12 Taler.*
  - *If in this second a white ball is drawn, you receive 12 Taler and your payoff after allocating your endowment is 52 Taler.*
- *If in the first random draw the ball "Down" is drawn, you receive 0 Taler. After allocating your endowment of 80 Taler for this decision to the lottery outcome your payoff is 80 Taler.*

Stage III

In stage III you are asked to make 10 decisions on a single decision screen. The risky events between you have to decide in this stage are displayed in a table format. In each row of the table you make one decision. For an illustration see the following figure:



Remaining time [sec]: 1799

	Option A				Option B				
Decision	Chance of Outcome 1	Outcome 1 (in Taler)	Chance of Outcome 2	Outcome 2 (in Taler)	Chance of Outcome 1	Outcome 1 (in Taler)	Chance of Outcome 2	Outcome 2 (in Taler)	Your decision
25	10%	20.00	90%	16.00	10%	38.50	90%	1.00	Option A ○ ○ Option B
26	20%	20.00	80%	16.00	20%	38.50	80%	1.00	Option A ○ ○ Option B
27	30%	20.00	70%	16.00	30%	38.50	70%	1.00	Option A ○ ○ Option B
28	40%	20.00	60%	16.00	40%	38.50	60%	1.00	Option A ○ ○ Option B
29	50%	20.00	50%	16.00	50%	38.50	50%	1.00	Option A ○ ○ Option B
30	60%	20.00	40%	16.00	60%	38.50	40%	1.00	Option A ○ ○ Option B
31	70%	20.00	30%	16.00	70%	38.50	30%	1.00	Option A ○ ○ Option B
32	80%	20.00	20%	16.00	80%	38.50	20%	1.00	Option A ○ ○ Option B
33	90%	20.00	10%	16.00	90%	38.50	10%	1.00	Option A ○ ○ Option B
34	100%	20.00	0%	16.00	100%	38.50	0%	1.00	Option A ○ ○ Option B

OK

Each risky event comprises two possible outcomes and two corresponding probabilities. You make your decision at the end of each row by indicating the risky event you prefer (either Option A or Option B). When making your decisions you do not have to follow a particular order and you can change your decisions as often as desired within the time permitted.

The outcomes of the risky events in this stage do not comprise losses. Thus, for the decisions in this stage you do not receive an endowment. Accordingly, your payoff is as follows:

Outcome of the risky event

How is the outcome of the chosen risky event determined in Stage III? To determine the outcome there is a ballot box with 100 balls marked with numbers from 1 to 100 (analogously to stage I). Each number occurs exactly once in the ballot box, i.e., the draw of a particular number is equally likely.

Before the experiment will start now, please note: You are asked comprehension questions before each stage starts. These questions should familiarize you with the decision task in each stage.

After the experiment, you are asked to answer a questionnaire. For answering the questionnaire you receive independently from your earnings during the experiment €4.

A.2 Appendix to Chapter 2

A.2.1 Derivations

We first show that our lottery preference B_2 over A_2 implies risk aversion in the differentiable EU framework, i.e., $u'' < 0$. B_2 is preferred to A_2 by an EU maximizer implies that

$$\begin{aligned} \frac{1}{2}u(x-k) + \frac{1}{2}u(x-r) &> \frac{1}{2}u(x-r-k) + \frac{1}{2}u(x) \\ \iff u(x) - u(x-k) &< u(x-r) - u(x-r-k). \end{aligned}$$

Now we divide by k and since the preference holds for all positive k we can let k go to zero to obtain

$$u'(x) < u'(x-r).$$

Since this holds for all r (and for all x) the latter equation implies that $u'(x)$ is strictly decreasing, i.e., $u''(x) < 0$ what we wanted to show. That the preferences B_3 over A_3 and B_4 over A_4, respectively, are equivalent to prudence and temperance within the differentiable EU framework is proven by use of similar arguments in Eeckhoudt and Schlesinger (2006).

We now present the approximations that relate individuals' indices of absolute risk attitude to the risk premia measured in our experiment.[8] The 2nd-degree risk premium that makes the individual indifferent between the risk loving and risk averse lottery choice is defined as

$$\frac{1}{2}u(x-k) + \frac{1}{2}u(x-r) = \frac{1}{2}u(x+m^{\mathrm{RA}}) + \frac{1}{2}u(x-r-k+m^{\mathrm{RA}}). \tag{6}$$

We approximate

$$\begin{aligned} u(x-k-r+m^{\mathrm{RA}}) &\approx u(x-r) + (m^{\mathrm{RA}}-k)u'(x-r) \\ &\approx u(x-r) + (m^{\mathrm{RA}}-k)\left(u'(x) - ru''(x)\right) \end{aligned}$$

[8] These approximations are similar to those in Crainich and Eeckhoudt (2008) who note that they are à la Arrow-Pratt.

such that equation (6) approximately becomes

$$\begin{aligned} u(x) - kU'(x) + u(x-r) &= u(x) + m^{\text{RA}}u'(x) + U(x-r) + (m^{\text{RA}} - k)\left(u'(x) - ru''(x)\right) \\ \Longleftrightarrow 0 &= m^{\text{RA}}u'(x) + m^{\text{RA}}\left(u'(x) - ru''(x)\right) + rku''(x) \\ \Longleftrightarrow 0 &= 2m^{\text{RA}}u'(x) + u''(x)\left((r(k - m^{\text{RA}})\right) \\ \Longleftrightarrow -\frac{u''(x)}{u'(x)} &= \frac{2m^{\text{RA}}}{r(k - m^{\text{RA}})}. \end{aligned}$$

For prudence, consider

$$\frac{1}{2}u(x-k) + \frac{1}{2}u(x+\epsilon) = \frac{1}{2}u(x + m^{\text{PR}}) + \frac{1}{2}u(x - k + \epsilon + m^{\text{PR}}). \tag{7}$$

We approximate

$$\begin{aligned} u(x - k + \epsilon + m^{\text{PR}}) &\approx u(x - k + m^{\text{PR}}) + \frac{1}{2}u''(x - k + m^{\text{PR}})\sigma^2 \\ &\approx u(x) + (m^{\text{PR}} - k)u'(x) + \frac{\sigma^2}{2}(u''(x) + (m^{\text{PR}} - k)u'''(x)) \end{aligned}$$

such that equation (7) approximately becomes

$$\begin{aligned} 2u(x) - ku'(x) + \frac{1}{2}\sigma^2 u''(x) &= 2u(x) + m^{\text{PR}}u'(x) + (m^{\text{PR}} - k)u'(x) \\ &\quad + \frac{\sigma^2}{2}\left(u''(x) + (m^{\text{PR}} - k)u'''(x)\right) \\ \Longleftrightarrow 0 &= 2m^{\text{PR}}u'(x) + \frac{\sigma^2}{2}(m^{\text{PR}} - k)u'''(x) \\ \Longleftrightarrow \frac{u'''(x)}{u'(x)} &= \frac{4m^{\text{PR}}}{\sigma^2(k - m^{\text{PR}})}. \end{aligned}$$

Finally, for temperance first consider

$$u(x + m^{\text{TE}} + \epsilon_1 + \epsilon_2)$$

which is approximated as

$$\begin{aligned} & u(x + \epsilon_1 + \epsilon_2) + m^{\text{TE}}u'(x + \epsilon_1 + \epsilon_2) \\ &= u(x + \epsilon_1 + \epsilon_2) + m^{\text{TE}}(u'(x) + u''(x)E[(\epsilon_1 + \epsilon_2)] \\ &= u(x + \epsilon_1 + \epsilon_2) + m^{\text{TE}}u'(x) \\ &= u(x + \epsilon_1) + \frac{1}{2}u''(x + \epsilon_1)\sigma_2^2 \\ &= u(x) + \frac{1}{2}u''(x)\sigma_1^2 + \frac{\sigma_2^2}{2}\left(u''(x) + \frac{1}{2}u^{(4)}(x)\sigma_1^2\right) \\ &= u(x) + \frac{1}{2}u''(x)\sigma_1^2 + \frac{1}{2}u''(x)\sigma_2^2 + \frac{1}{4}u^{(4)}(x)\sigma_1^2\sigma_2^2 \end{aligned}$$

Thus we can approximate

$$u(x+\epsilon_1)+u(x+\epsilon_2) \quad = \quad u(x+m^{\mathrm{TE}})+u(x+m^{\mathrm{TE}}+\epsilon_1+\epsilon_2) \tag{8}$$

as

$$u(x)+u''(x)\sigma_1^2+u(x)+u''(x)\sigma_2^2 = u(x)+m^{\mathrm{TE}}u'(x)+u(x)+\frac{1}{2}u''(x)\sigma_1^2+\frac{1}{2}u''(x)\sigma_2^2+\frac{1}{4}u^{(4)}(x)\sigma_1^2\sigma_2^2.$$

Collecting terms and rearranging yields

$$\begin{aligned} 0 &= 2m^{\mathrm{TE}}u'(x)+\frac{1}{4}u^{(4)}(x)\sigma_1^2\sigma_2^2 \\ \Longleftrightarrow m^{\mathrm{TE}} &= -\frac{u^{(4)}(x)}{u'(x)}\cdot\left(\frac{1}{8}\sigma_1^2\sigma_2^2\right). \end{aligned}$$

A.2.2 Instructions

[*translated from German*]

Thank you very much for participating in this decision experiment!

General Information

In the following experiment, you will make a couple of decisions. Following the instructions and depending on your decisions, you can earn money. It is therefore very important that you read the instructions carefully.

You will make your decisions anonymously on your computer screen in your cubicle. During the experiment you are not allowed to talk to the other participants. Whenever you have a question, please raise your hand. The experimenter will answer your question in private in your cubicle. If you disregard these rules you can be excluded from the experiment. Then you receive no payment.

During the experiment **all amounts are stated in Euro**. At the end of the experiment, your achieved earnings will be paid to you in cash.

Structure of the Experiment

The experiment can be divided into three stages. All stages are equally relevant for your payoff. The three stages comprise decision problems, where risky events play a role. In a risky event it is unsure, which outcome occurs.

You decide, which of two risky events you prefer. The form of the risky events will be described when explaining the stages in-depth.

Overall you will make 120 individual decisions in the three sections of the experiment.

Payoff in the experiment

To determine your payoff of the experiment, one of your 120 decisions from the three sections will be selected randomly. This takes place after you have made all your decisions. For this the experimenter will draw one out of 120 cards, labeled with numbers from 1 to 120, from a ballot-box. Every number occurs only once in the ballot-box whereby the draw of a particular number is equally likely. The outcome of the risky event, that you have chosen will actually be determined afterwards. These random draws will be explained in-depth when describing the sections of the experiments.

Note that only one of your 120 decisions determines your earnings of the experiment and that each of your 120 decisions can determine your entire earnings of the experiment.

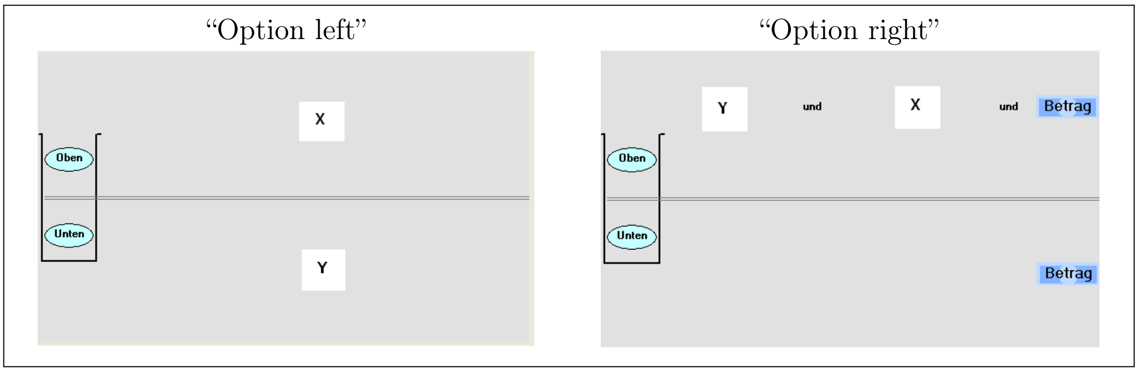
Also note that the risky events can comprise negative outcomes. You receive an endowment in form of a coupon. The coupons are allocated to the outcome of the risky event. Hence your payoff is made up of the two components

Endowment and **Outcome of the chosen risky event**.

After the experiment, the decision relevant for payoff and the outcome of the risky event will be randomly will be determined for each participant in the seminar room. For this participants will be called on successively.

Decision situation

The risky events displayed in following figure describe the decision situation, you face in the three stages of the experiment, in an abstract way. In decision situation you decide which of the two risky events (here: “Option left” and “Option right”) you prefer.



Both the risky event “Option left” and “Option right” comprise one random draw (Random draw 1), that is depicted by the balls “Up” and “Down”. Random draw 1 is: With 50% chance you are in state “Up” or with 50% chance in state “Down”.

We now look at the risky event “**Option left**”: If the ball “Up” will be drawn, the outcome is X. X can either be a fixed amount or another random draw (Random draw X). If ball “Down” is drawn, the outcome is Y. Also Y can either be a fixed amount or another random draw (Random draw Y).

In risky event “**Option right**” X and Y follow, if Ball “Up” is drawn. In addition a Amount (blue bank note) is added to both state “Up” and state “Down”. If ball “Down” is drawn, you receive the amount indicated on the bank note. If ball “Up” is drawn, X and Y follow and the Amount (blue bank note) is added.

The Amount on the blue bank note can take the following values

$$\boxed{\mathbf{-2.50}, \mathbf{-2.25}, \mathbf{-2.00}, \ldots, \mathbf{-0.25}, \mathbf{0.00}, \mathbf{0.25}, \ldots, \mathbf{2.00}, \mathbf{2.25}.}$$

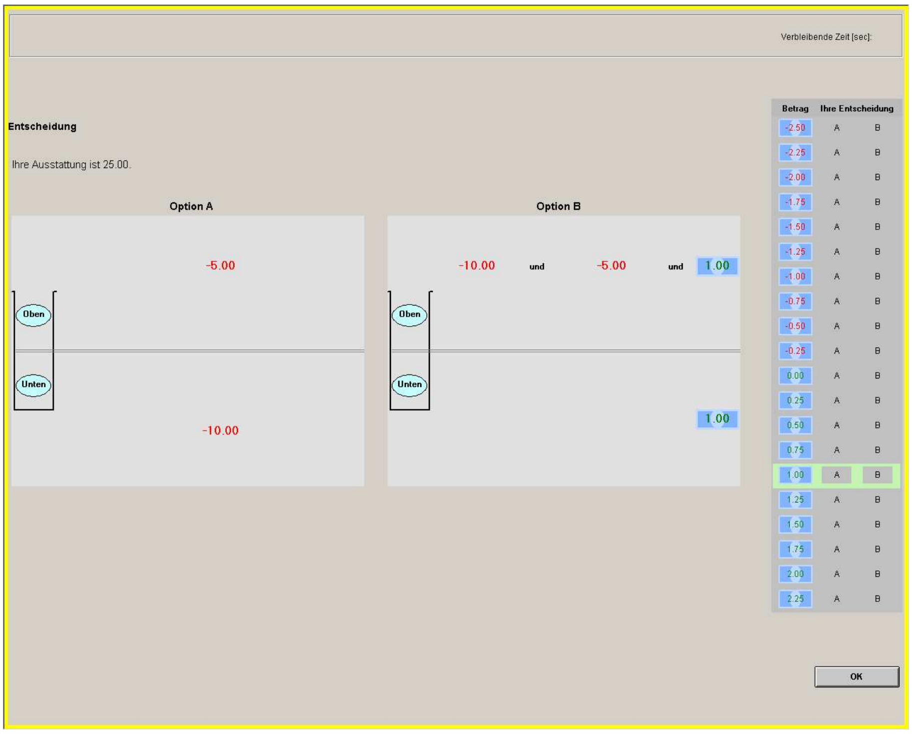
Hence, for each of these 20 Amounts follows one decision situation with two risky events. The Amount on the blue bank note is always added to the states “Up” and “Down” of that risky event, where both X and Y occur in state “Up” (here: “Option right”).

Note that, on your decision screen on the computer the risky event, where the Amount (blue bank note) is added can either be the right or the left option.

First stage

In the first stage of the experiment you make 20 decisions. You choose on one decision screen at a time, which of the two different risky events—Option A or Option B—you prefer.

The risky events can comprise negative outcomes. For each decision in the first stage you receive an endowment of 25.00. An example of a decision situation in the first stage is provided in the following figure.



In the example above the Amount (blue bank note) is added to Option B. The size of the added Amount can be found in the column "Amount" on the right-hand side of the screen. For each Amount you decide whether you prefer Option A or Option B.

After activating an Amount in the column "Amount" you decide for this Amount by clicking on "A" or "B" whether you prefer Option A or Option B. A green frame marks the chosen option. You do not need to stick to a certain order of your decisions.

How is the outcome of the risky event (you have chosen) determined in the first stage? For Random draw 1 there are two balls in a ballot-box—one with the label "Up" another with the label "Down". Both balls can be drawn with the same chance.

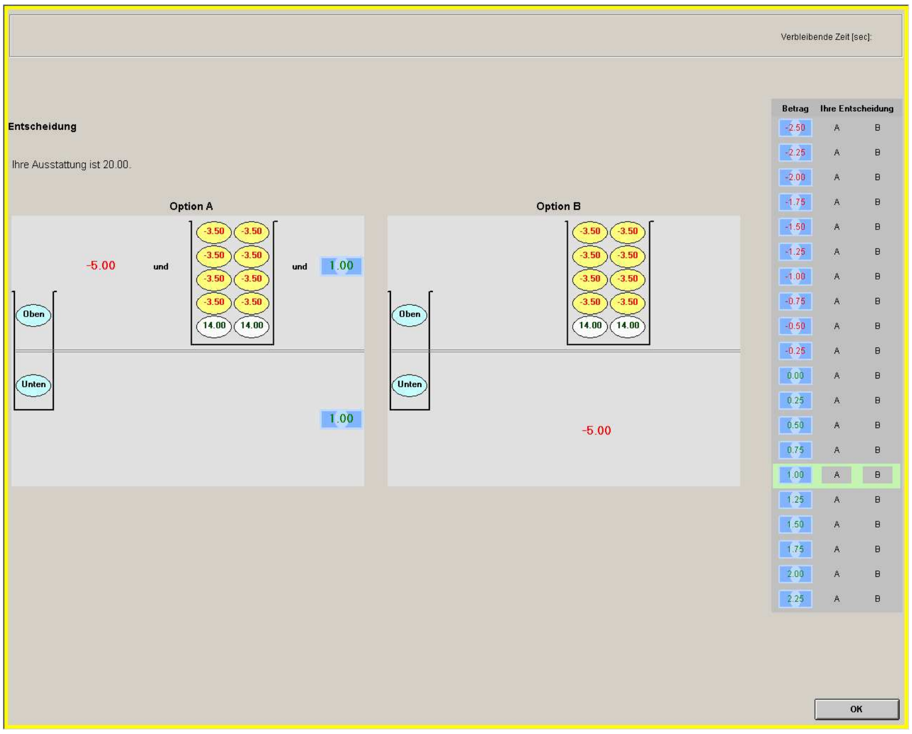
Please look at the example of this stage again!
Suppose, this decision has been randomly chosen to determine your payoff. In **Option A** *the outcome is* −5.00*, if in* Random draw 1 *the ball* "Up" *is drawn. If the ball* "Down" *is drawn the outcome is* −10.00*. Considering your* Endowment *of* 25.00 *in Option A results* "Up" 20.00 *and in stage* "Down" 15.00.

In **Option B** *the outcome is* −10.00 *and* −5.00 *and* 1.00 (Amount *on the blue bank* *note), if in* Random draw 1 "Up" *is drawn; overall* −14.00*. If ball* "Down" *is drawn, the outcome is* 1.00 (Amount on the blue bank note)*. Considering your* Endowment *of* 25.00 *in Option B results in stage* "Up" 11.00 *and in stage* "Down" 26.00 *for your payoff.*

Second stage

In the second stage of the experiment you make <u>60</u> decisions. You choose on three decision screens each with 20 decision situtations, which of the two different risky events—Option A or Option B—you prefer.

The outcomes of the risky events can be negative. You receive an Endowment of <u>20.00</u>. An example of a decision situation in the second stage is provided in the following figure.



In the example above the Amount (blue bank note) is added to Option A. The size of the added Amount can be found in the column "Amount" on the right-hand side of the screen. For <u>each</u> Amount you decide whether you prefer Option A or Option B.

After activating an Amount in the column "Amount" you decide for this Amount by clicking on "A" or "B" whether you prefer Option A or Option B. A green frame marks the chosen option. You do not need to stick to a certain order of your decisions.

How is the outcome of the risky event (you have chosen) determined in the second stage? For Random draw 1 there are two balls in a ballot-box—one with the label "Up" another with the label "Down". Both balls can be drawn with the same chance (analogous to the first stage). As shown in the example above, in the second stage a second random draw (RANDOM DRAW X) can be necessary to determine your payoff. In RANDOM DRAW X a ball is drawn from a ballot-box containing 10 balls. This ball can either be white or yellow. Note that, the composition of white and yellow balls can change in the three decision screens in this stage. This ballot-box always contains 10 balls and within a decision screen (for 20 decisions) the composition of white and yellow balls are identical.

Please look at the example of this stage again!

Suppose, this decision situation has been randomly chosen to determine your payoff. If in **Option A** *in* RANDOM DRAW 1 *the ball* "**Up**" *is drawn, the outcome is* −5.00, RANDOM DRAW X *follows and* 1.00 (AMOUNT *on the blue bank note).*

- *If in* RANDOM DRAW X *a* <u>*yellow*</u> *ball is drawn, you lose* 3.50. *Considering your* ENDOWMENT *of* 20.00 *you receive* 12.50 *(*$= 20.00 - 5.00 - 3.50 + 1.00$*).*
- *If in* RANDOM DRAW X *a* <u>*white*</u> *ball is drawn, you receive* 14.00. *Considering your* ENDOWMENT *you receive* 30.00 *(*$= 20.00 - 5.00 + 14.00 + 1.00$*).*

If in **Option A** *in* RANDOM DRAW 1 "Down" *is drawn, the outcome is* 1.00 *(*AMOUNT *on the blue bank note). Considering your* ENDOWMENT 21.00 *result.*

If in **Option B** *in* RANDOM DRAW 1 "Up" *is drawn,* RANDOM DRAW X *follows.*

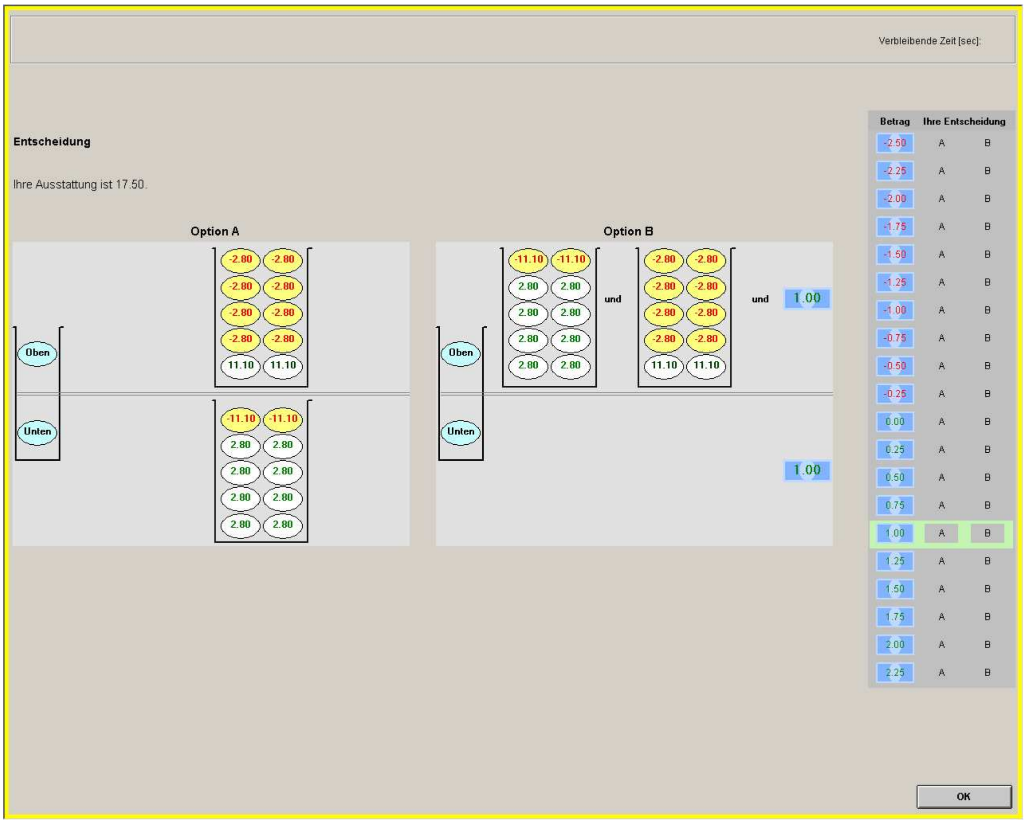
- *If in* RANDOM DRAW X *a* <u>*yellow*</u> *ball is drawn, you lose* 3.50. *Considering your* ENDOWMENT *of* 20.00 *you receive* 16.50.
- *If in* RANDOM DRAW X *a* <u>*white*</u> *ball is drawn, you receive* 14.00. *Considering your* ENDOWMENT *you receive* 34.00.

If in **Option B** *in* RANDOM DRAW 1 "Down" *is drawn, the outcome is* −5.00. *Considering your* ENDOWMENT 15.00 *result.*

Third stage

In the second stage of the experiment you make <u>40</u> decisions. You choose on two decision screens each with 20 decision situations, which of the two different risky events—Option A or Option B—you prefer.

The outcomes of the risky events can be negative. You receive an ENDOWMENT of <u>17.50</u>. An example of a decision situation in the third stage is provided in the following figure.



In the example above the Amount (blue bank note) is added to Option B. The size of the added Amount can be found in the column "Amount" on the right-hand side of the screen. For each Amount you decide whether you prefer Option A or Option B.

After activating an Amount in the column "Amount" you decide for this Amount by clicking on "A" or "B" whether you prefer Option A or Option B. A green frame marks the chosen option. You do not need to stick to a certain order of your decisions.

How is the outcome of the risky event (you have chosen) determined in the third stage? For Random draw 1 there are two balls in a ballot-box—one with the label "Up" another with the label "Down". Both balls can be drawn with the same chance (analogous to the first and second stage).

As shown in the example above, in the second stage a second random draw (Random draw X) and/or a third random draw (Random draw Y) can be necessary to determine your payoff.

In Random draw X a ball is drawn from a ballot-box containing 10 balls. This ball can either be white or yellow. Note that, the composition of white and yellow balls can change in the three decision screens in this stage. This ballot-box always contains 10 balls and within a decision screen (for 20 decisions) the composition of white and yellow balls are identical. Analogously, this is true for Random draw Y. Notice that the composition of yellow and white balls across Random draw X and Random draw Y can differ (see the example above).

Please look at the example of this stage again!
Suppose, this decision situation has been randomly chosen to determine your payoff.

If in **Option A** *in* Random draw 1 *the ball* "Up" *is drawn,* Random draw X *follows.*

- *If in* Random draw X *a yellow ball is drawn, you lose* 2.80. *Considering your* Endowment *of* 17.50 *you receive* 14.70.
- *If in* Random draw X *a white ball is drawn, you receive* 11.10. *Considering your* Endowment *you receive* 28.60.

If in Random draw 1 *the ball* "Down" *is drawn,* Random draw Y *follows.*

- *If in* Random draw Y *a yellow ball is drawn, you lose* 11.10. *Considering your* Endowment *of* 17.50 *you receive* 6.40.
- *If in* Random draw Y *a white ball is drawn, you receive* 2.80. *Considering your* Endowment *you receive* 20.30.

If in **Option B** *in* Random draw 1 "Up" *is drawn* Random draw Y *and* Random draw X *follow and the* Amount *of* 1.00 *(blue bank note) is added.*

- *If in* Random draw X *and in* Random draw Y *a yellow ball is drawn, you lose* 2.80 *(from* Random draw X *) and* 11.10 *(from* Random draw Y*). Considering your* Endowment *of* 17.50 *you receive* 4.60 *(*$= 17.50 - 11.10 - 2.80 + 1.00$*).*
- *If in* Random draw X and in Random draw Y a white ball is drawn, you receive 11.10 (from Random draw X) and 2.80 (from Random draw Y). *Considering your* Endowment *of* 17.50 *you receive* 32.40 *(*$= 17.50 + 11.10 + 2.80 + 1.00$*).*
- *If in* Random draw X *a white ball and in* Random draw Y *a yellow ball is drawn, you receive* 11.10 *(from* Random draw X) *and you lose* 11.10 *(from* Random draw Y*). Considering your* Endowment *you receive* 18.50 *(*$= 17.50 + 11.10 - 11.10 + 1.00$*).*

- *If in* Random draw X *a* <u>*yellow*</u> *ball and in* Random draw Y *a* <u>*white*</u> *ball is drawn, you lose* 2.80 *(from* Random draw X*) and you receive* 2.80 *(from* Random draw Y*). Considering your* Endowment *you receive* 18.50 *(*$= 17.50 - 2.80 + 2.80 + 1.00$*).*

If in **Option B** *in* Random draw 1 "Down" *is drawn, the outcome is* 1.00 (Amount *on the blue bank note). Considering your* Endowment 18.50 *result.*

Before the experiment will start now, please note: You are asked comprehension questions before each stage starts. These questions should familiarize you with the decision task in each stage.

After the experiment, you are asked to answer a questionnaire. For answering the questionnaire you receive independently from your earnings during the experiment € 4.

A.3 Appendix to Chapter 3

A.3.1 Tables and figures

Table A.3.1: Quantities and differences to patient optimal quantity by physician

	Physician i	q_{ij} Mean	q_{ij} Median	q_{ij} s.d.	μ_{ij} Mean	μ_{ij} Median	μ_{ij} s.d.
FFS	1	6.40	7.00	1.12	1.40	1.00	1.35
	2	7.73	8.00	1.87	2.73	2.00	2.94
	3	5.00	5.00	1.46	0.00	0.00	0.65
	4	5.00	5.00	1.69	0.00	0.00	0.00
	5	7.27	8.00	1.16	2.27	2.00	1.49
	6	6.40	6.00	1.12	1.40	1.00	1.80
	7	7.13	7.00	1.06	2.13	2.00	1.77
	8	8.27	9.00	1.94	3.27	3.00	2.69
	9	6.07	7.00	1.39	1.07	1.00	1.28
	10	7.67	7.00	1.76	2.67	2.00	2.55
	11	7.47	8.00	2.00	2.47	2.00	2.61
	12	6.93	7.00	1.75	1.93	2.00	2.05
	13	6.13	6.00	1.92	1.13	1.00	1.92
	14	6.27	7.00	1.33	1.27	1.00	1.62
	15	8.53	9.00	1.96	3.53	3.00	2.85
	16	6.67	6.00	1.54	1.67	1.00	2.47
	17	5.00	5.00	1.69	0.00	0.00	0.00
	18	5.73	6.00	1.49	0.73	1.00	1.03
	19	7.00	7.00	1.25	2.00	2.00	1.96
	20	5.33	5.00	1.45	0.33	0.00	1.11
CAP	1	4.20	5.00	1.52	-0.80	-1.00	1.61
	2	4.27	5.00	0.88	-0.73	0.00	1.28
	3	4.80	5.00	1.47	-0.20	0.00	0.41
	4	5.13	5.00	1.60	0.13	0.00	0.52
	5	2.13	2.00	0.83	-2.87	-4.00	1.46
	6	5.00	5.00	1.69	0.00	0.00	0.00
	7	4.07	4.00	0.96	-0.93	-1.00	1.10
	8	4.33	5.00	0.98	-0.67	0.00	0.98
	9	4.07	4.00	0.80	-0.93	-1.00	1.22
	10	5.00	5.00	1.69	0.00	0.00	0.00
	11	4.93	5.00	1.62	-0.07	0.00	0.26
	12	4.93	5.00	1.62	-0.07	0.00	0.26
	13	2.40	2.00	1.18	-2.60	-2.00	2.38
	14	5.00	5.00	1.69	0.00	0.00	0.00
	15	4.00	4.00	0.85	-1.00	-1.00	0.85
	16	4.47	5.00	1.85	-0.53	0.00	2.00
	17	3.40	4.00	1.68	-1.60	-3.00	1.99
	18	4.53	5.00	1.19	-0.47	0.00	0.74
	19	6.00	6.00	2.45	1.00	2.00	2.56
	20	4.67	5.00	1.29	-0.33	0.00	0.49
	21	5.00	5.00	1.69	0.00	0.00	0.00
	22	4.47	5.00	1.13	-0.53	0.00	0.83

This table shows descriptive statistics of quantities for each individual physician (columns 2-4). Further, it shows descriptive statistics about the differences of the actual quantity to the patient optimal quantity, i.e., $\mu_{ij} = q_{ij} - q_j^*$ (columns 5-7).

Table A.3.2: Average quantities and ranks by patient type

		Average quantity						Number of
	Physician i	Pat. type 1	Pat. type 2	Pat. type 3	Order of ranks			inversions
FFS	1	6.4	5.6	7.2	2	1	3	0
	2	8.0	8.4	6.8	2	3	1	1
	3	5.0	3.4	6.6	2	1	3	0
	4	5.0	3.0	7.0	2	1	3	0
	5	7.2	6.6	8.0	2	1	3	0
	6	5.8	6.4	7.0	1	2	3	1
	7	7.2	6.8	7.4	2	1	3	0
	8	8.4	8.4	8.0	2	2	1	.
	9	6.8	4.6	6.8	2	1	2	.
	10	8.2	7.6	7.2	3	2	1	2
	11	6.8	7.8	7.8	1	2	2	.
	12	7.6	6.0	7.2	3	1	2	3
	13	6.4	5.0	7.0	2	1	3	0
	14	5.8	5.8	7.2	1	1	2	.
	15	8.6	9.0	8.0	2	3	1	1
	16	6.6	7.0	6.4	2	3	1	1
	17	5.0	3.0	7.0	2	1	3	0
	18	6.0	4.2	7.0	2	1	3	0
	19	7.0	6.8	7.2	2	1	3	0
	20	5.8	3.8	6.4	2	1	3	0
CAP	1	3.6	3.6	5.4	2	2	1	.
	2	4.6	3.4	4.8	2	1	3	0
	3	5.0	3.0	6.4	2	1	3	0
	4	5.0	3.4	7.0	2	1	3	0
	5	1.4	2.0	3.0	1	2	3	1
	6	5.0	3.0	7.0	2	1	3	0
	7	4.4	3.0	4.8	2	1	3	0
	8	5.0	3.0	5.0	2	1	2	.
	9	4.4	3.2	4.6	2	1	3	0
	10	5.0	3.0	7.0	2	1	3	0
	11	5.0	3.0	6.8	2	1	3	0
	12	5.0	3.0	6.8	2	1	3	0
	13	3.0	2.6	1.6	3	2	1	2
	14	5.0	3.0	7.0	2	1	3	0
	15	4.0	3.0	5.0	2	1	3	0
	16	5.0	3.4	5.0	2	1	2	.
	17	3.4	2.8	4.0	2	1	3	0
	18	5.0	3.0	5.6	2	1	3	0
	19	6.4	5.0	6.6	2	1	3	0
	20	5.0	3.0	6.0	2	1	3	0
	21	5.0	3.0	7.0	2	1	3	0
	22	5.0	3.0	5.4	2	1	3	0

This table shows physicians' average quantity choices by patient type (columns 2-4) and the corresponding ranks (5-7). The final column shows the number of inversions necessary to reach the 'perfect order of ranks', i.e., 2,1,3, starting from the observed order of ranks.

Table A.3.3: Test statistics of two-sided non-parametric tests per patient

		Decision j (Patient)														
	Test; Variable(s); Scope	1 (1A)	2 (1B)	3 (1C)	4 (1D)	5 (1E)	6 (2A)	7 (2B)	8 (2C)	9 (2D)	10 (2E)	11 (3A)	12 (3B)	13 (3C)	14 (3D)	15 (3E)
I	Binomial; num. q_j^*, num. not q_j^*; FFS	0.0000	0.0414	0.1153	0.0118	0.0025	0.0414	0.0004	0.0414	0.0025	0.5034	0.5034	0.0414	1.0000	0.5034	0.0118
	Binomial; num. q_j^*, num not. q_j^*; CAP	0.1338	0.1338	0.1338	0.5235	0.1338	0.2863	0.0043	0.0001	0.0001	0.0001	0.5235	0.1338	0.2863	0.2863	0.0524
II	Mann Whitney U; q_{ij}; across treatments	0.2440	0.0000	0.0000	0.0000	0.0000	0.0010	0.0000	0.0000	0.0000	0.0001	0.2339	0.0000	0.0002	0.0001	0.0000
III	Fisher exact; q_{ij}^*; across treatments	0.0063	0.0051	0.0000	0.0095	0.0005	0.0111	0.0000	0.0000	0.0000	0.0006	0.2461	0.0051	0.1670	0.0784	0.2457
IV	Mann Whitney U; $\psi(q_{ij})$; across treatments	0.0066	0.1539	0.3047	0.4630	0.0617	0.0271	0.0000	0.0000	0.0000	0.0008	0.3972	0.0028	0.0507	0.1206	0.5991
V	Mann Whitney U; $\Delta\pi_{ij}$; across treatments	0.0000	0.0195	0.8055	0.0590	0.2046	0.0183	0.0008	0.2092	0.0119	0.0203	0.0000	0.0002	0.0000	0.0000	0.4219

This table shows test statistics by decision j (patient). Row I provides p-values of a two-sided binomial test comparing the number of patient optimal quantity choices with the number of patient non-optimal quantity choices for FFS and CAP. In row II, p-values of a two-sided Mann-Whitney U test are shown comparing physicians' quantity choices across treatments. Row III gives p-values of a Fischer exact test about the number of patient optimal quantity choices, i.e., q_{ij}^*, across treatments. p-values of a two-sided Mann-Whitney U test comparing benefit losses $\psi(q_{ij})$ and relative deviations from the physician optimal profit $\Delta\pi_{ij}$ are shown in rows IV and V, respectively.

Table A.3.4: Descriptive statistics on variables per patient

			Decision j (Patient)														
			1 (1A)	2 (1B)	3 (1C)	4 (1D)	5 (1E)	6 (2A)	7 (2B)	8 (2C)	9 (2D)	10 (2E)	11 (3A)	12 (3B)	13 (3C)	14 (3D)	15 (3E)
FFS	q_{ij}	Mean ($\overline{q}_j$)	5.00	7.30	6.40	6.80	7.90	4.70	6.20	5.85	6.60	6.45	6.30	7.10	7.35	7.35	7.70
		Median	5.00	7.00	6.50	6.80	8.00	5.00	7.00	6.00	6.00	8.00	6.00	7.00	7.00	7.00	8.00
	ν_{ij}	Mean ($\overline{\nu}_j$)	0.00	2.30	1.40	1.80	2.90	1.70	3.20	2.85	3.60	3.45	-0.70	0.10	0.35	0.35	0.70
		Median	0.00	2.00	1.50	1.00	3.00	2.00	4.00	3.00	3.00	5.00	-1.00	0.00	0.00	0.00	1.00
		s.d.	0.00	1.84	1.50	1.67	1.86	1.22	2.24	2.21	2.23	3.05	0.86	0.64	0.75	0.88	1.03
	$\pi(q_{ij})$	Mean	8.00	10.18	7.21	10.79	10.29	6.98	9.03	6.68	10.88	9.18	7.43	10.66	7.78	12.20	10.28
		Median	8.00	11.10	7.45	11.40	11.60	8.00	11.10	7.20	11.40	11.60	7.40	11.10	7.70	12.00	11.60
		s.d.	0.00	2.60	0.90	2.82	3.04	1.66	3.26	1.41	2.92	3.62	0.31	1.68	0.25	0.51	2.20
	$\Delta\pi_{ij}$	Mean	0.00	0.19	0.12	0.21	0.21	0.13	0.28	0.20	0.20	0.29	0.07	0.15	0.06	0.10	0.21
		Median	0.00	0.11	0.09	0.16	0.11	0.00	0.11	0.13	0.16	0.11	0.08	0.11	0.07	0.12	0.11
		s.d.	0.00	0.21	0.11	0.21	0.23	0.21	0.26	0.17	0.21	0.28	0.04	0.13	0.03	0.04	0.17
	$B(q_{ij})$	Mean	10.00	8.85	8.85	9.10	8.55	9.15	8.40	8.58	8.20	8.28	8.92	9.21	9.04	8.90	8.47
		Median	10.00	9.00	9.00	9.50	8.50	9.00	8.00	8.50	8.50	7.50	9.00	9.45	9.23	9.45	8.80
	$\psi(q_{ij})$	Mean ($\overline{\psi}_j$)	0.00	1.15	1.15	0.90	1.45	0.85	1.60	1.43	1.80	1.73	0.53	0.25	0.41	0.55	0.98
		Median	0.00	1.00	1.00	0.50	1.50	1.00	2.00	1.50	1.50	2.50	0.45	0.00	0.23	0.00	0.65
		s.d.	0.00	0.92	1.73	0.84	0.93	0.61	1.12	1.10	1.12	1.53	0.64	0.62	0.62	0.96	1.48
		obs.	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
CAP	q_{ij}	Mean ($\overline{q}_j$)	4.73	4.59	4.27	4.64	4.55	3.45	3.18	3.05	2.86	3.00	5.59	5.50	5.77	5.64	5.18
		Median	5.00	5.00	5.00	5.00	5.00	3.00	3.00	3.00	3.00	3.00	6.00	5.50	5.50	6.00	5.50
	ν_{ij}	Mean ($\overline{\nu}_j$)	-0.27	-0.41	-0.73	-0.36	-0.45	0.45	0.18	0.05	-0.14	0.00	-1.41	-1.50	-1.23	-1.36	-1.82
		Median	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-1.00	-1.50	-1.50	-1.00	-1.50
		s.d.	0.98	1.18	1.28	1.81	1.34	1.37	0.85	0.49	0.47	0.93	1.74	1.41	1.69	1.43	1.82
	π_{ij}	Mean	9.67	9.76	10.02	9.54	9.76	10.63	10.92	11.05	11.16	11.02	8.59	8.79	8.40	8.63	9.00
		Median	9.50	9.50	9.50	9.50	9.50	11.10	11.10	11.10	11.10	11.10	8.40	8.95	8.95	8.40	8.95
		s.d.	0.72	0.92	0.86	1.87	0.93	1.29	0.60	0.36	0.20	0.62	1.52	1.42	2.02	1.45	1.55
	$\Delta\pi_{ij}$	Mean	0.19	0.19	0.17	0.21	0.19	0.11	0.09	0.08	0.07	0.08	0.28	0.27	0.30	0.28	0.25
		Median	0.21	0.21	0.21	0.21	0.21	0.08	0.08	0.08	0.08	0.08	0.30	0.25	0.25	0.30	0.25
		s.d.	0.06	0.08	0.07	0.16	0.08	0.11	0.05	0.03	0.02	0.05	0.13	0.12	0.17	0.12	0.13
	$B(q_{ij})$	Mean	8.99	8.60	8.01	8.31	8.57	8.91	9.45	9.57	9.20	9.48	7.99	7.94	7.77	8.07	7.49
		Median	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	9.00	8.38	7.75	9.00	8.38
	$\psi(q_{ij})$	Mean ($\overline{\psi}_j$)	1.01	1.40	1.99	1.69	1.43	1.09	0.55	0.43	0.80	0.52	1.46	1.51	1.68	1.38	1.96
		Median	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.45	1.08	1.70	0.45	1.08
		s.d.	2.18	2.69	3.40	2.88	2.80	2.55	1.92	1.81	2.58	2.14	2.30	1.91	2.14	1.94	2.56
		obs.	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22

Table A.3.5: Descriptive statistics on patient benefit and profit by physician

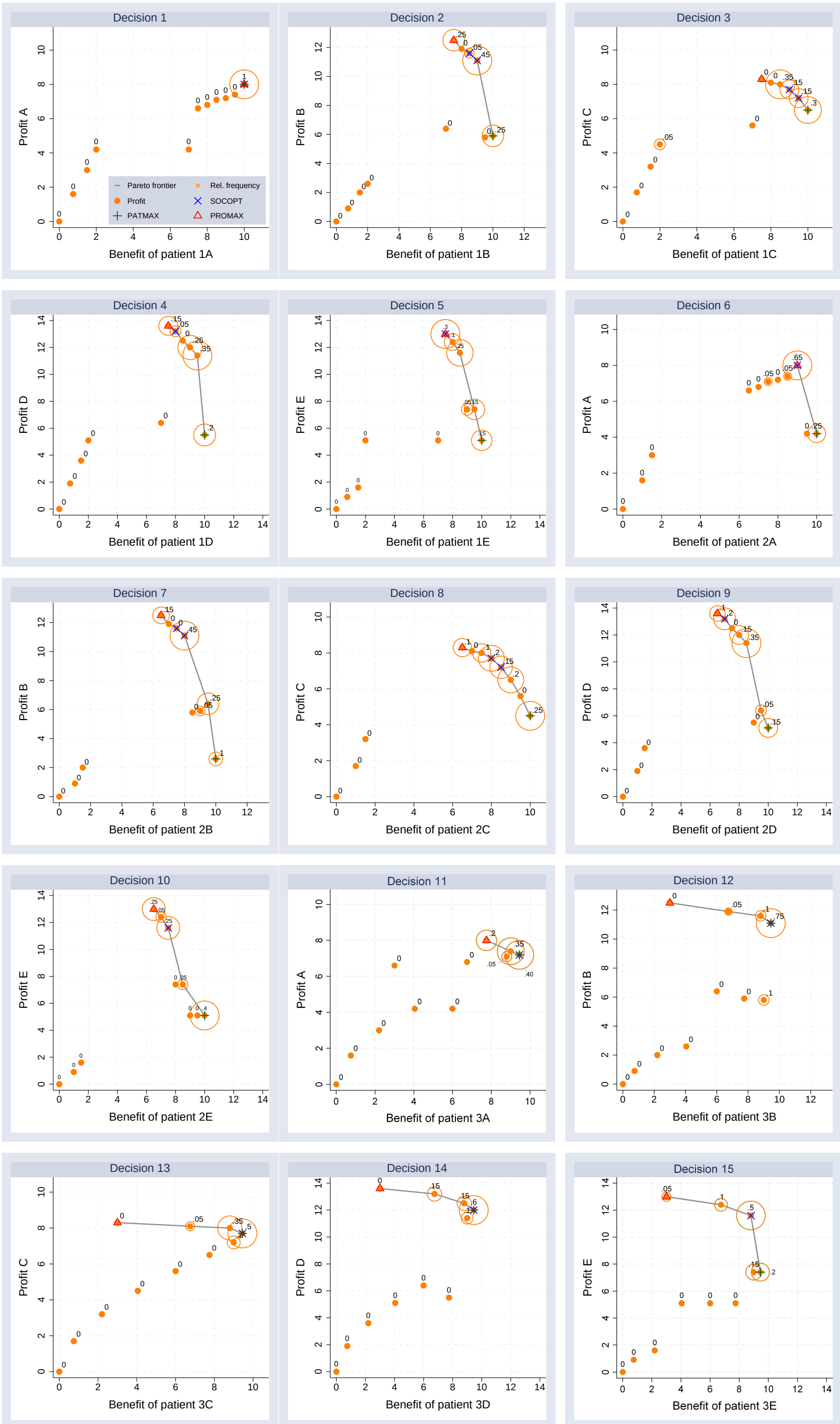
	FFS						CAP					
	Patient benefit $B(q_{ij})$			Profit $\pi(q_{ij})$			Patient benefit $B(q_{ij})$			Profit $\pi(q_{ij})$		
Physician i	Mean	Median	s.d.	Mean	Median	s.d.	Mean	Median	s.d.	Mean	Median	s.d.
1	9.11	9.45	0.55	8.83	8.00	1.93	8.25	9.00	2.54	10.02	9.50	1.03
2	8.19	8.50	1.08	10.09	11.60	2.79	8.67	9.00	1.43	10.11	9.50	0.71
3	9.69	10.00	0.42	6.73	5.90	2.27	9.73	10.00	0.42	9.49	9.50	1.41
4	9.82	10.00	0.27	6.53	5.90	2.51	9.75	10.00	0.34	9.13	9.50	1.62
5	8.63	8.80	0.59	10.15	11.10	2.11	2.73	1.50	2.45	11.48	11.60	0.34
6	9.04	9.00	0.75	9.50	11.10	2.26	9.82	10.00	0.27	10.26	10.40	0.82
7	8.66	8.80	0.63	10.15	11.10	1.98	8.50	9.00	1.61	11.48	11.60	0.34
8	7.54	7.75	1.63	10.81	12.00	2.38	9.25	10.00	1.10	9.23	9.50	1.70
9	9.22	9.00	0.45	8.88	8.00	2.37	8.38	7.75	1.59	10.03	9.50	0.78
10	8.47	8.80	1.10	10.46	11.10	2.50	9.82	10.00	0.27	10.29	10.40	0.64
11	7.68	7.50	1.90	10.24	11.10	2.76	9.79	10.00	0.33	9.23	9.50	1.70
12	8.84	9.00	0.89	9.81	11.10	2.64	9.79	10.00	0.33	9.32	9.50	1.62
13	9.18	9.45	0.86	8.99	8.00	2.90	4.87	2.20	3.91	11.29	11.60	0.55
14	9.17	9.45	0.73	9.37	11.10	2.41	9.82	10.00	0.27	9.23	9.50	1.70
15	7.52	7.50	1.10	10.93	12.40	2.48	8.25	7.75	1.32	10.33	10.40	0.68
16	8.79	9.00	0.95	9.70	11.10	2.30	8.74	9.50	2.58	9.69	9.50	1.48
17	9.82	10.00	0.27	6.53	5.90	2.51	6.47	6.00	3.54	10.58	10.40	0.97
18	9.38	9.45	0.44	8.69	8.00	2.52	9.50	10.00	0.82	9.81	9.50	1.03
19	8.51	8.50	0.88	10.11	11.10	1.79	7.23	8.50	2.90	7.84	8.40	2.90
20	9.39	9.00	0.49	6.87	6.50	1.68	9.67	10.00	0.49	9.67	9.50	1.15
21	.	.	.	.	.	.	9.82	10.00	0.27	9.23	9.50	1.70
22	.	.	.	.	.	.	9.42	10.00	0.93	9.89	9.50	0.96

Table A.3.6: Relative frequencies of choices on the Pareto frontier sorted by categories

		Decision j (Patient)														
	Category	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
		(1A)	(1B)	(1C)	(1D)	(1E)	(2A)	(2B)	(2C)	(2D)	(2E)	(3A)	(3B)	(3C)	(3D)	(3E)
FFS	PROMAX	1.00‡	0.25	0.00	0.15	0.32‡	0.72‡	0.16	0.10	0.10	0.25	0.21	0.00	0.00	0.00	0.06
FFS	PATMAX	1.00‡	0.25	0.32	0.20	0.16	0.28	0.11	0.25	0.15	0.40	0.42‡	0.83‡	0.56‡	0.67‡	0.24
FFS	SOCOPT	1.00‡	0.50	0.32	0.05	0.32‡	0.72‡	0.47	0.35	0.20	0.25	0.42‡	0.83‡	0.56‡	0.00‡	0.59
FFS	PAROTH	1.00‡	0.00	0.37	0.60	0.53	0.00	0.26	0.30	0.55	0.10	0.37	0.17	0.44	0.33	0.71
CAP	PROMAX	0.00	0.00	0.00	0.05	0.05	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.00	0.00	0.05
CAP	PATMAX	0.75‡	0.71‡	0.68‡	0.65‡	0.71‡	0.88‡	0.95‡	0.95‡	0.91‡	0.95‡	0.41	0.32	0.38	0.36	0.27
CAP	SOCOPT	0.75‡	0.71‡	0.68‡	0.65‡	0.71‡	0.88‡	0.95‡	0.95‡	0.91‡	0.95‡	0.18	0.18	0.10	0.23	0.23
CAP	PAROTH	0.25	0.29	0.32	0.30	0.24	0.13	0.05	0.05	0.09	0.00	0.36	0.50	0.52	0.41	0.45

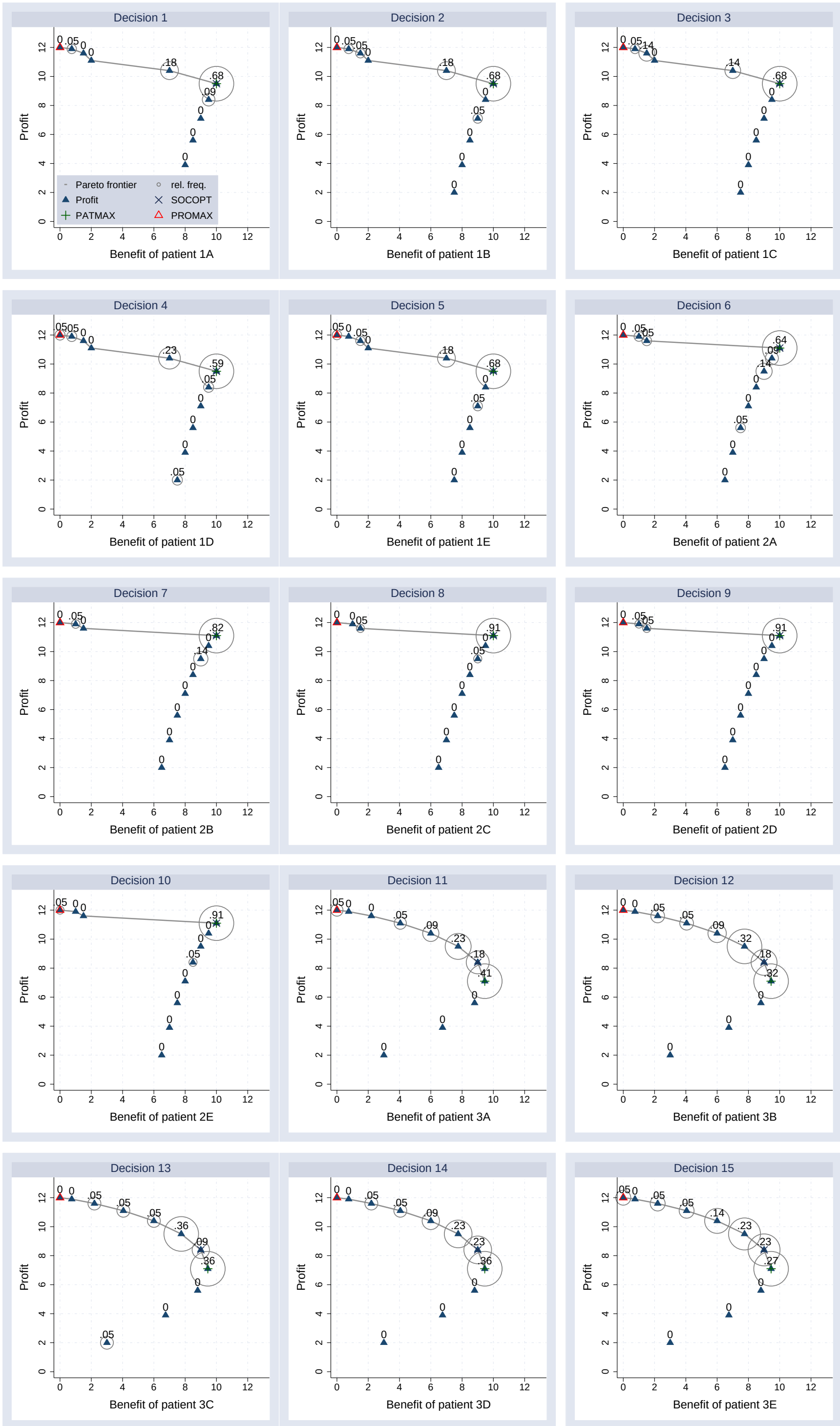
‡ Some patient benefit/profit pairs are covered by two (or more) *different* categories. In particular, the social optimal quantity (q_j^{soc}) coincides for some decisions with either the profit maximal quantity ($\hat{q}_j$) or the patient optimal quantity (q_j^*).
In FFS: $(\pi(\hat{q}_j), B(\hat{q}_j)) = (\pi(q_j^*), B(q_j^*)) = (\pi(q_j^{soc}), B(q_j^{soc}))$ for patient 1A ($j = 1$); $(\pi(\hat{q}_j), B(\hat{q}_j)) = (\pi(q_j^{soc}), B(q_j^{soc}))$ for patients 1E and 2A ($j = 5, 6$); $(\pi(q_j^*), B(q_j^*)) = (\pi(q_j^{soc}), B(q_j^{soc}))$ for for patients 3A,..,3D (j =11,..,14).
In CAP: $(\pi(q_j^*), B(q_j^*)) = (\pi(q_j^{soc}), B(q_j^{soc}))$ for for patients 1A,...,2E ($j = 1, \ldots, 10$).

Figure A.3.1: Pareto frontiers per decision in FFS



This figure contrasts the patient benefit (vertical axis) and the physician profit (horizontal axis) for each decision in FFS. Further, it illustrates the Pareto efficient benefit/profit pairs on the Pareto frontier, the patient optimal choice (symbol: +), the social optimal choice (×) and the profit maximizing choice for the physician (Δ). Circles indicate relative frequencies of physicians' choices for each benefit/profit pair.

Figure A.3.2: Pareto frontiers per decision in CAP



This figure contrasts the patient benefit (vertical axis) and the physician profit (horizontal axis) for each decision in CAP. Further, it illustrates the Pareto efficient benefit/profit pairs on the Pareto frontier, the patient optimal choice (symbol: +), the social optimal choice (×) and the profit maximizing choice for the physician (Δ). Circles indicate relative frequencies of physicians' choices for each benefit/profit pair.

A.3.2 Instructions

[*translated from German*]

General Information

In the following experiment, you will make a couple of decisions. Following the instructions and depending on your decisions, you can earn money. It is therefore very important to read the instructions carefully.

You take your decisions anonymously in your cubicle on your computer screen. During the experiment you are not allowed to talk to any other participant. Whenever you have a question, please raise your hand. The experimenter will answer your question in private in your cubicle. If you disregard these rules you can be excluded from the experiment without receiving any payment.

All amounts in the experiment are stated in Taler. After the experiment, your earnings will be converted at an exchange rate of 1 Taler = 0.05 EUR and paid to you in cash.

Your decisions in the experiment

During the entire experiment you are in the role of a physician. You have to decide on the treatment of 15 patients. All participants of this experiment are taking their decisions in the role of a physician. You decide on the **quantity** of medical services you want to provide for a given illness of a patient.

You decide on your computer screen where five different illnesses, A, B, C, D and E, of three different patient types, 1, 2 and 3, will be shown one after another. For each patient you can provide $0, 1, 2, 3, 4, 5, 6, 7, 8, 9$ or 10 medical services. Your remuneration is as follows:

- *Treatment FFS*:
  A different PAYMENT is assigned to each **quantity** of medical services . The PAYMENT increases in the **quantity** of medical services.
- *Treatment CAP*:
  For each patient you receive a lump-sum PAYMENT that is independent of the **quantity** of medical services.

While deciding on the **quantity** of medical services, in addition to your PAYMENT you determine the COSTS you incur when providing these services. COSTS increase with increasing **quantity** provided. Your PROFIT in Taler is calculated by subtracting your COSTS from your PAYMENT.

To each quantity of medical services a certain benefit for the patient is assigned, the PATIENT BENEFIT that the patient gains from your provision of services (treatment). Therefore, your decision on the **quantity** of medical services not only determines your own PROFIT but also the PATIENT BENEFIT. An example for a decision situation is given on the following screen.

Screen shot FFS

Illness E / Patient 1

Medical services	Quantity	Your Payment (in Taler)	Your Cost (in Taler)	Your Profit (in Taler)	Patient benefit (in Taler)
none	0	0.00	0.00	0.00	0.00
Service E1	1	1.00	0.10	0.90	0.75
Service E1, Service E2	2	2.00	0.40	1.60	1.50
Service E1, Service E2, Service E3	3	6.00	0.90	5.10	2.00
Service E1, Service E2, Service E3, Service E4	4	6.70	1.60	5.10	7.00
Service E1, Service E2, Service E3, Service E4, Service E5	5	7.60	2.50	5.10	10.00
Service E1, Service E2, Service E3, Service E4, Service E5 Service E6	6	11.00	3.60	7.40	9.50
Service E1, Service E2, Service E3, Service E4, Service E5 Service E6, Service E7	7	12.30	4.90	7.40	9.00
Service E1, Service E2, Service E3, Service E4, Service E5 Service E6, Service E7, Service E8	8	18.00	6.40	11.60	8.50
Service E1, Service E2, Service E3, Service E4, Service E5 Service E6, Service E7, Service E8, Service E9	9	20.50	8.10	12.40	8.00
Service E1, Service E2, Service E3, Service E4, Service E5 Service E6, Service E7, Service E8, Service E9, Service E10	10	23.00	10.00	13.00	7.50

Please indicate the quantity of medical services you wish to provide:

OK

Screen shot CAP

Illness E / Patient 1

Medical services	Quantity	Your Payment (in Taler)	Your Cost (in Taler)	Your Profit (in Taler)	Patient benefit (in Taler)
none	0	12.00	0.00	12.00	0.00
Service E1	1	12.00	0.10	11.90	0.75
Service E1, Service E2	2	12.00	0.40	11.60	1.50
Service E1, Service E2, Service E3	3	12.00	0.90	11.10	2.00
Service E1, Service E2, Service E3, Service E4	4	12.00	1.60	10.40	7.00
Service E1, Service E2, Service E3, Service E4, Service E5	5	12.00	2.50	9.50	10.00
Service E1, Service E2, Service E3, Service E4, Service E5 Service E6	6	12.00	3.60	8.40	9.50
Service E1, Service E2, Service E3, Service E4, Service E5 Service E6, Service E7	7	12.00	4.90	7.10	9.00
Service E1, Service E2, Service E3, Service E4, Service E5 Service E6, Service E7, Service E8	8	12.00	6.40	5.60	8.50
Service E1, Service E2, Service E3, Service E4, Service E5 Service E6, Service E7, Service E8, Service E9	9	12.00	8.10	3.90	8.00
Service E1, Service E2, Service E3, Service E4, Service E5 Service E6, Service E7, Service E8, Service E9, Service E10	10	12.00	10.00	2.00	7.50

Please indicate the quantity of medical services you wish to provide:

OK

You decide on the **quantity** of medical services on your computer screen by typing an integer between 0 and 10 into the box named "Your Decision".

There are no real but abstract patients participating in this experiment. Yet the PATIENT BENEFIT an abstract patient receives by your providing medical services will be beneficial for a real patient. The total amount corresponding to the sum over all 15 PATIENT BENEFITS determined by your decisions will be transferred to the charity *Christoffel Blindenmission Deutschland e.V., 64625 Bensheim*, to support an ophthalmic hospital where patients with cataract are treated.

Earnings in the experiment

After having made your 15 decisions, your overall earnings will be calculated by summing up the PROFITs from all your decisions. This amount will be converted from Taler into Euro at the end of the experiment.

The overall PATIENT BENEFIT resulting from your 15 quantity decisions will be converted into Euro as well and will be transferred to the *Christoffel Blindenmission*.

The transferral will be made by the experimenter and a monitor. The monitor writes a check on the amount of money corresponding to the aggregated PATIENT BENEFITS of this experiment. This check issued to the *Christoffel Blindenmission* will be sealed in an envelope addressed to this charity. The monitor and experimenter then walk together to the nearest mailbox and deposit the envelope.

After all participants have taken their decisions, one participant is randomly assigned the role of the monitor. The monitor receives a payment of 4 EUR in addition to the payment from the experiment. The monitor verifies, by a signed statement, that the procedure described above was actually carried out.

Next, please answer some questions familiarizing you with the decision situation.

After your 15 decisions, please answer some further questions on your screen.

A.4 Appendix to Chapter 4

Figure A.4.1: Proportional sacrifices per decision and subject in Med-FFS

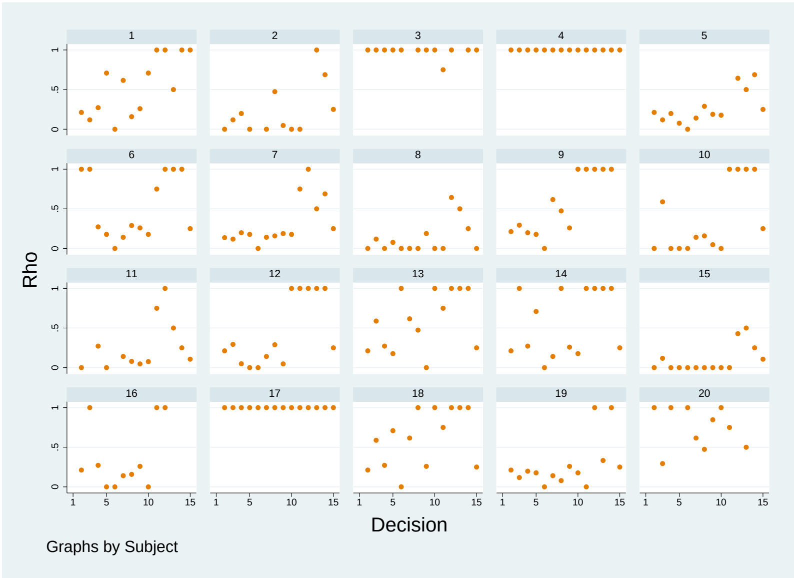


Figure A.4.2: Proportional sacrifices per decision and subject in Non-Med-FFS

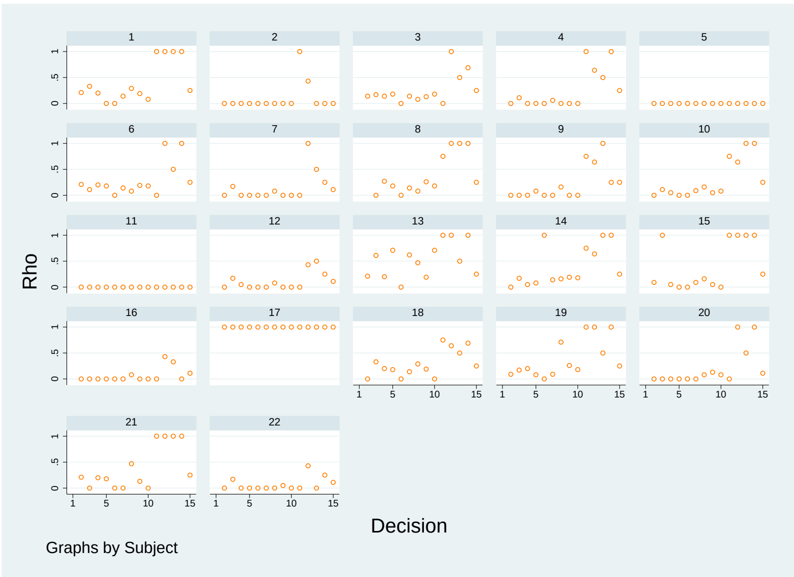


Figure A.4.3: Proportional sacrifices per decision and subject in Med-CAP

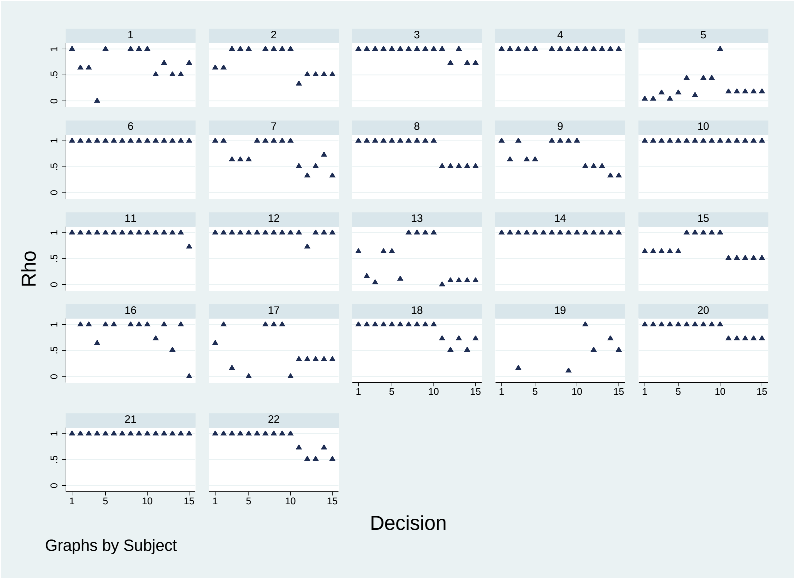


Figure A.4.4: Proportional sacrifices per decision and subject in Non-Med-CAP