# Deductibles and Health Care Utilization: An Experiment on the Role of Forward-Looking Behavior 

Glenn W. Harrison*, Nadja Kairies-Schwarz ${ }^{\dagger}$ and Johann Han ${ }^{\dagger \ddagger}$<br>15 February 2024


#### Abstract

We investigate the effects of nonlinear deductible contracts on health utilization behavior by using a laboratory experiment in which we can control the likelihood of hitting the deductible. We also evaluate the effect of subjects receiving regular information updates on their remaining deductible. Our results show that varying the future price has a significant effect on health care consumption. At an individual level, we identify and richly characterize heterogeneity. We find fully forward-looking, fully myopic, as well as mixed types after controlling for risk preferences. We show that there is a substantial welfare loss due to a lack in forward-looking behavior. The distribution and drivers of the welfare loss are characterized, and differ sharply according to the model of risk preferences adopted for normative evaluation.


JEL classification: I13, C91, D6, D82, D9.
Keywords: Health insurance, nonlinear prices, forward-looking behavior, laboratory experiment, risk preferences, welfare evaluation.

[^0]
## 1. Introduction

In an effort to reduce health care spending, policy makers, health insurance companies, and employers have tried to incorporate some form of cost sharing in their policies. One popular way to do this is to include deductibles in health care plans. There is also an efficiency rationale for including deductibles in insurance contracts, that derives from a moral hazard problem when the insured has no incentive to engage in low-cost effort to mitigate risks. A deductible aligns the interests of the insured and the insurer, up to a point. ${ }^{1}$ Such deductible plans affect health care prices dynamically over a period of time since an individual will initially pay higher prices for the same flow of health services before reaching the deductible than after reaching the deductible.

How health insurance pricing influences consumers' medical spending has been actively studied with field data. The first major contribution was the RAND Health Insurance Experiment, followed by the Oregon Health Insurance Experiment. ${ }^{2}$ Given the nonlinear pricing structure over time that is central to deductible plans, how do consumers react to implied future prices as well as to the current spot price? Recent studies have studied the effects of nonlinear contracts on utilization behavior either using firm or claim level data for health insurance plans, or data for Medicare Part D plans. ${ }^{3}$

To investigate the effect of dynamic incentives created by deductible plans on health care utilization behavior, one might keep the spot price constant while generating variation in the future price. For instance, Aron-Dine et al. (2015) use an empirical strategy that compares individuals who join the same deductible health plans at different times of the year. These individuals face the same spot price, when they join, but different future prices due to variations in the remaining time until the end of the year when the deductible resets. Using claim-level data from employer-provided health insurance in the U.S., they conclude that consumers indeed react to more than just the spot price. Initial health care utilization is higher for individuals who joined early in the year and thus face a low future price. However, their approach depends on the assumption that reasons for joining in different months can be viewed as exogenous to health care utilization behavior.

[^1]Irrespective of the empirical strategy, each of the previously mentioned field studies investigating the effects of nonlinear contracts on utilization behavior face potential confounders, which make it difficult to manipulate the future price while holding the spot price constant. While seasonality might be relatively easy to account for, liquidity constraints, intertemporal substitution and comorbidities might substantially affect utilization and are difficult to control for in the field.

We complement the empirical field work on the effects of nonlinear deductible contracts on health utilization behavior by using a controlled laboratory experiment. Although there has been a growing literature on health economic experiments involving health insurance ${ }^{4}$, we are the first to address this research question with a laboratory experiment. Compared to the field, the laboratory allows one to perfectly control for a constant spot price while varying the future price, as well as for other confounding factors. In the experiment, subjects go through a cycle of periods and are insured by a health plan with a deductible. In each period they face probabilistic health events and have to choose between seeking treatment or not. Similar to Harrison and Ng (2016) and Kairies-Schwarz et al. (2017), we also elicit risk preferences from each subject and infer risk preferences at the level of the individual. This allows to derive individually optimal treatment choices that we then compare to their actual decisions, allowing us to normatively evaluate the welfare effects of nonlinear deductible contracts.

We investigate various factors that might influence health care utilization in the context of dynamic incentives. First, we manipulate the channel through which the same future price is generated by exogenously varying the contract duration length or the deductible level. Second, we exogenously vary whether subjects receive regular information updates on their accumulated costs and remaining deductible. This information may be relevant in the context of episodic healthcare utilization over time. Health insurance plans with deductibles are complex, and it is not clear how well individuals understand their insurance policies. ${ }^{5}$ Not understanding their health insurance plans implies that individuals may not respond correctly to the incentives. One way to improve understanding of health care plans and thus health care utilization is to provide individuals with better information about their health plans, or simplify the decision process. ${ }^{6}$

[^2]In line with the empirical evidence investigating the effects of nonlinear contracts on health utilization behavior, we find that individuals do respond to the dynamic incentives created by deductible plans. Our results show that the future price has a significant effect on spending behavior. The channel by which the future price is manipulated, whether the same future price is reached by changing the deductible or the number of periods, seems to be secondary. Controlling for individual risk preferences, we find fully forward-looking individuals as well as fully myopic individuals.

It is one thing to identify in detail the behavioral effects of deductibles, and related informational treatments, and another thing to show that these effects lead to a welfare gain for individuals. Perhaps some individuals made mistakes when processing the choice tasks before them. To evaluate the welfare gains or losses from changes in observed behavior, we also provide a structural evaluation of latent effects on expected consumer surplus. We show that individuals insured with an insurance contract with a deductible face substantial welfare losses due to a lack in forward-looking behavior, irrespective of whether we assume a mixture of Rank Dependent Expected Utility (RDU) risk and Expected Utility Theory (EUT) preferences or EUT preferences. Under the mixture of EUT and RDU and EUT we see a large number of small welfare losses and a long tail of larger losses. The number of the latter is slightly higher for EUT than for the mixture of EUT and RDU along the tail. These results point to more individuals making mistakes that were welfare costly under EUT. Finally, we show that the drivers of these welfare effects also differ depending on which model of risk preferences is used for normative evaluation. Under the mixture of EUT and RDU, for example, we find a significant welfare reduction for women, while under EUT there is a significant increase in welfare for women.

In Section 2 we lay out the experimental design and procedures. In Section 3 we report our results. In Section 4 we show limitations before presenting conclusions in Section 4.
for low income families choosing schools with high test scores, Bhargava and Manoli (2015) for economically and socially disadvantaged families claiming eligible tax benefits, or Bhargava et al. (2017) and Samek and Sydnor (2020) for health plan choices with dominated options.

## 2. Experimental Design

### 2.1 Decision Situation

## Basic Decision Scenario

We employed a laboratory experiment with a sequential design. In the first part, we elicited individual risk preferences. In the second part we analyzed health utilization behavior under dynamic incentives. ${ }^{7}$

The design of the risk preference elicitation in the first part was similar to Andersen et al. (2008) where subjects made decisions over a battery of binary choices over risky lotteries. Each subject made 20 decisions to identify risk preferences (See Appendix A.1.1).

In the second part of the experiment, we investigated subjects' health care utilization behavior given dynamic incentives. This part varied between treatment conditions as shown in Table $1 .{ }^{8}$ In a periodic task, subjects went through a cycle of decision situations. Each period a subject received an income of 50 ECU (experimental currency unit) ${ }^{9}$ and faced one of three possible events: (a) healthy, (b) sickness A or (c) sickness B. The health events were drawn from a probability distribution known to all participants. The 'healthy' event occurred with probability 0.6 , and 'sickness A' and 'sickness B' each occurred with probability 0.2.

Conditional on the realization of the health event, subjects had to choose their action. If a subject was healthy, they did not face any costs, did not make any decision, kept the periodic income, and moved on to the next period. If a subject was sick, they had to decide whether to get treated or not. Costs for both decisions did not differ for sickness A ( 50 ECU each), but it was cheaper to leave the sickness untreated with sickness B (30 ECU) than getting treatment (again 50 ECU ). Hence, in the absence of insurance, sickness A could be interpreted as the relatively more severe sickness. The optimal decision for the relatively less severe sickness B would be no treatment in a one shot game in which the decision to seek treatment occurs after realizing that it is sickness B. ${ }^{10}$ There were no consequences of the decision on future health outcomes or probabilities, and subjects knew that.

[^3]Table 1: Potential Health States, Costs of (Not) Treating, and Event Probabilities

| Health State | Cost of Choosing <br> Treatment | Cost of Not Choosing <br> Treatment | Probability of Event |
| :--- | :---: | :---: | :---: |
| Healthy | 0 | 0 | 0.6 |
| Sickness A | 50 | 50 | 0.2 |
| Sickness B | 50 | 30 | 0.2 |

Notes. This table shows the potential health states, and their respective costs of treating and not treating, as well as the event probabilities.

Depending on the condition, subjects went through 52 or 26 decision periods. The order of health events was drawn prior to the experiment and was the same between all treatment conditions at least until the 26th period, since some sessions ended after that. Hence, ex ante, the spot price of care is constant across individuals. Table 2 displays the order of exogenous health events that subjects faced. In contrast to the field there is no need to control different histories of health events or "seasonality", where locally or temporarily concentrated health events could affect utilization behavior.

Table 2: Draw of Health Events for Every Period

| Period | $\mathbf{1}$ | $\mathbf{2}$ | $\mathbf{3}$ | $\mathbf{4}$ | $\mathbf{5}$ | $\mathbf{6}$ | $\mathbf{7}$ | $\mathbf{8}$ | $\mathbf{9}$ | $\mathbf{1 0}$ | $\mathbf{1 1}$ | $\mathbf{1 2}$ | $\mathbf{1 3}$ |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Health | G | B | G | G | A | G | G | B | G | G | A | G | A |
|  | $\mathbf{1 4}$ | $\mathbf{1 5}$ | $\mathbf{1 6}$ | $\mathbf{1 7}$ | $\mathbf{1 8}$ | $\mathbf{1 9}$ | $\mathbf{2 0}$ | $\mathbf{2 1}$ | $\mathbf{2 2}$ | $\mathbf{2 3}$ | $\mathbf{2 4}$ | $\mathbf{2 5}$ | $\mathbf{2 6}$ |
|  | G | B | G | B | B | G | G | G | B | G | G | G | A |
|  | $\mathbf{2 7}$ | $\mathbf{2 8}$ | $\mathbf{2 9}$ | $\mathbf{3 0}$ | $\mathbf{3 1}$ | $\mathbf{3 2}$ | $\mathbf{3 3}$ | $\mathbf{3 4}$ | $\mathbf{3 5}$ | $\mathbf{3 6}$ | $\mathbf{3 7}$ | $\mathbf{3 8}$ | $\mathbf{3 9}$ |
|  | B | G | G | B | A | A | G | B | G | G | B | B | G |
|  | $\mathbf{4 0}$ | $\mathbf{4 1}$ | $\mathbf{4 2}$ | $\mathbf{4 3}$ | $\mathbf{4 4}$ | $\mathbf{4 5}$ | $\mathbf{4 6}$ | $\mathbf{4 7}$ | $\mathbf{4 8}$ | $\mathbf{4 9}$ | $\mathbf{5 0}$ | $\mathbf{5 1}$ | $\mathbf{5 2}$ |
|  | A | G | G | A | G | G | G | B | G | G | B | G | A |

Notes. This table shows the draw of health events for every period, where $\mathrm{G}=\operatorname{good}$ health; $\mathrm{A}=$ sickness $\mathrm{A} ; \mathrm{B}=$ sickness B. The order of health events was drawn before the experiments and was identical for all sessions and conditions.

## Insurance

Apart from receiving the periodic income of 50 ECU , subjects were told that they had insurance coverage for negative health events after they had incurred medical expenditures beyond a deductible. The deductible level varied by the experimental condition. If subjects decided to treat a sickness, they paid the cost themselves up to the remaining deductible. Medical spending beyond the deductible was free.

In our benchmark condition LowPrice, the deductible was set at 600 ECU . In this case, subjects would need to pay for 12 treatment decisions and would have free treatment thereafter. The costs of leaving the sickness untreated, however, did not affect the deductible,
and subjects would always bear those costs out of their (cumulated) periodic income. In the case of sickness B, the treatment costs of 50 ECU would count towards the deductible, and the opportunity costs of non-treatment, 30 ECU , would not affect the deductible. Given the insurance setting, the optimal decision with sickness B becomes more interesting. After every period, subjects received information about their accumulated income and, depending on the condition, about their accumulated treatment costs and the remaining deductible.

The expected end-of-year price, or future price, $p^{f_{i t}}$ in our design plays a critical role in understanding the dynamic incentives central to our design. It lies between 0 and 1 and is
 period $t$ will hit the deductible by the end of the experiment if the subject treats all further sicknesses, regardless of the severity. For a health insurance plan with no deductible the $\operatorname{Pr}\left(h_{i t}\right)$ is naturally 1 and hence $p_{i t}^{f}$ is 0 . The lower $\operatorname{Pr}\left(h_{i t}\right)$ is the higher is $p_{i t}^{f}$. When $\operatorname{Pr}\left(h_{i t}\right)$ is 0 the future price $p^{f_{i t}}$ is 1 . To have a better comparison to the constant treatment spot price of 50 ECU and the opportunity cost of non-treatment of 30 for sickness B in the experiment, we then derive the normalized future price by multiplying $p^{f}{ }_{i t}$ by 50 . Hence, the normalized future price is 50 when the $\operatorname{Pr}\left(h_{i t}\right)$ is 0 and $p_{i}^{f}$ is 1 . The higher $\operatorname{Pr}\left(h_{i t}\right)$ is, the lower are the $p_{i}^{f}$ and the normalized future price.

The probability of hitting the deductible depends on the probability of falling sick, the level of the deductible, and the number of periods left in the game. ${ }^{11}$ These parameters can be exogenously manipulated in the experiment between conditions to create variation in the future price. Within any exogenous condition, the future price is identical for each subject at the beginning. We vary the future price compared to LowPrice through two channels: by reducing the number of periods from 52 to 26 while keeping the deductible of 600 constant (HighPricePeriod), and by increasing the deductible to 1150 while keeping the number of periods at 52.

### 2.2 Experimental Conditions

We conducted seven experimental conditions shown in Table 3. Part A of Table 3 shows our main treatment conditions. The objective of the conditions LowPrice, HighPricePeriods and HighPriceDed was to investigate the relationship between future price and expenditure by keeping everything constant other than the future price. In our benchmark condition LowPrice

[^4]the deductible was relatively low ( 600 ECU ) over the duration of 52 periods. The likelihood of spending beyond the deductible was therefore high, which implies a low future price of 0.003 translating into a normalized future price of 0.15 . The normalized future price is hence very low compared to the opportunity cost of not treating sickness B of 30. In HighPricePeriods we increased the future price by decreasing the number of periods to 26, and in HighPriceDed we increased the future price by increasing the deductible to 1150 ECU. Both manipulations led to similar (normalized) future prices, 0.674 (33.7) and 0.687 (34.35) respectively. The normalized future prices are hence above the opportunity cost of not treating sickness B (30).

Table 3: Experimental Conditions Overview

|  | Spot Price | Deductible | \# Periods | Normalized <br> Future <br> Price | $N$ | Add. Info |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
| Condition |  |  |  |  |  |  |
| Part A: Main Conditions |  | 600 | 52 | 0.15 | 48 | Yes |
| LowPrice | 50 | 600 | 26 | 33.7 | 24 | Yes |
| HighPricePeriods | 50 | 50 | 1150 | 52 | 34.35 | 24 |
| HighPriceDed | 50 | Yes |  |  |  |  |
| Part B: Robustness Conditions |  |  |  |  |  |  |
| LowPriceNoInfo | 50 | 600 | 52 | 0.15 | 48 | No |
| HighPriceNoInfo | 50 | 1150 | 52 | 34.35 | 24 | No |
| LowPriceReverse | 50 | 600 | 52 | 0.15 | 20 | Yes |
| LowPriceNeutral | 50 | 600 | 52 | 0.15 | 47 | Yes |
| Total |  |  |  |  | 235 |  |

Notes. This table shows the experimental conditions. In the main conditions we varied the level of the deductible as well as the number of periods. In the robustness conditions we checked for the effects of not giving information on the remaining level of the deductible in a HighPrice and a LowPrice scenario. We also checked for order effects when reversing the order of preference elicitation and healthcare consumption. Finally, we implemented a condition with a neutral insurance frame.

Part B of Table 3 shows our robustness checks. Conditions LowPriceNoInfo and HighPriceNoInfo aimed at investigating the role of information. Here subjects only learned their accumulated income, and not the accumulated treatment costs and remaining deductible, after each decision period. Condition LowPriceReverse was a control condition for order
effects, in which we started with the health utilization decisions followed by the elicitation of risk preferences. Condition LowPriceNeutral was an additional control in which the decision situation was framed in a neutral indemnity insurance context, with no association to health.

Since it has been shown that even small changes in wording can lead to significant effects on the results (see e.g. Levitt and List, 2007 for an overview), most recent health economic experiments, including those in the health insurance setting, use a health framing (see e.g. Kairies-Schwarz et al. 2017; Samek and Sydnor, 2020; Biener and Zou, 2022; Hermanns et al., 2023). However, insurance contracts with deductibles are also common in other insurance markets, and our results may be transferable to such other markets. Therefore, we implemented a condition with a neutral framing of the decision situation.

### 2.3 Behavioral Expectations

To derive behavioral expectations, we consider initial utilization, which we define as the number of treatments of sickness A and sickness B within the first 26 periods of the contract for two behavioral types (a fully forward-looking individual who takes all future periods into account and a myopic individual that does not look beyond the current period). For simplicity here, we assume risk neutrality.

The completely myopic individual only takes the costs of the current period into account. In the event of sickness B, this individual would never choose to treat for 50 ECU since the cost of non-treatment is only 30 ECU . In the event of sickness A , this individual would be indifferent between paying 50 for treating or incurring costs of 50 for non-treatment. ${ }^{12}$ Thus, the expected number of doctor visits for a myopic individual over the initial 26 periods would be 5.2 (or lower). ${ }^{13}$ Since a myopic individual does not consider the future price, this result holds regardless of the level of the deductible as well as the number of periods left to reach the deductible. In total contrast, for a hypothetical fully-insured individual, who would treat every sickness since this individual does not have to worry about cost, the expected number of doctor visits over the initial 26 periods would be 10.4. The completely forward-looking individual considers the future price. The expected initial utilization of this individual thus depends on the future price and differs between our HighPrice and LowPrice conditions. Initial utilization of a completely forward-looking individual in the LowPrice conditions with a deductible of 600 is comparable to a fully insured individual, since the likelihood of using

[^5]up the deductible is very high and the future price is close to $0 .{ }^{14}$ Increasing the deductible, while holding the remaining periods constant as in our HighPriceDed contract, increases the future price and reduces the expected initial utilization of the completely forward-looking individual. ${ }^{15}$

In order to derive specific behavioral expectations for our experimental conditions, we assume that individuals choose their individual treatment strategy based on the implied parameters before the utilization periods start and then stick to it. To build intuition, consider the two pure strategies of always treating only sickness A and always treating both sickness A and B. ${ }^{16}$ In our LowPrice conditions, always treating sickness A and B can then be considered the forward-looking strategy. Given the actual draw of sicknesses within the experiment, this strategy would yield 10 treatments over the first 26 periods. Treating sickness A only can be considered the myopic strategy in our LowPrice conditions resulting in 4 treatments. In the HighPrice conditions, both forward-looking and myopic types would choose the strategy to treat sickness A only ${ }^{17}$, resulting in 4 treatments over the first 26 periods. Hence we have these two hypotheses:

Hypothesis 1: Assuming risk neutrality and an ex ante choice of a pure decision strategy, forward-looking (myopic) individuals will choose the strategy to always treat sickness A and B (treat sickness A only) and will treat 10 (4) times in the first 26 periods in our LowPrice conditions. In the HighPrice conditions, both forward-looking and myopic individuals will choose to treat sickness A only and will treat 4 times in the first 26 periods.

Hypothesis 2: If the future price is relevant, we expect aggregate treatment rates to be higher in our LowPrice than in our HighPrice conditions. This difference should result from different treatment rates for sickness B.

To substantiate our hypotheses, we consider the profits generated by these two pure strategies for our LowPrice and HighPrice conditions when the actual draw of sicknesses within the experiment is taken into account. The upper panel A of Figure 1 displays the

[^6]profit ${ }^{18}$ outcomes for a deductible of 600 and thus our LowPrice and HighPricePeriod conditions. In the LowPrice conditions, the strategy to always treat sickness A and B leads to the deductible being hit at period 30. Subsequent treatment is then free of cost. Always treating Sickness A and B, leads to higher profits than only treating sickness A, and profits thereafter remain higher until period 52. In the HighPricePeriod condition, the experiment ends after 26 periods. At that point, the profit lines have not crossed yet and treating sickness A only is more profitable. Similarly, in the lower panel B of Figure 1, simulating the outcomes after period 26, we observe that, given a high future price due to a high deductible, HighPriceDed, always treating sickness A and B is less profitable than treating the A only.

Figure 1: Profit Outcomes by Treatment Strategy


Panel B: Simulated Outcomes for the HighPriceDed Strategy


[^7]
### 2.3 Experimental Procedure

The experiments were conducted at the Essen Laboratory for Experimental Economics (elfe) in Essen in 2016 and 2017. Sessions lasted about 90 minutes and 235 students participated after being recruited by the online recruiting system ORSEE due to Greiner (2015).

Subjects were randomly assigned to their seats in the laboratory upon arrival. Before each part of the experiment, they were given the appropriate instructions and were given time to read them and ask questions (see Appendix A.1.2, available on request). Any questions were answered in private by the same experimenter across all sessions. To ensure understanding of the decision task in each part, subjects had to answer a set of control questions (see Appendix A.1.3, available on request), and the experiment did not start unless all subjects had answered the control questions correctly.

At the end of the experiment, two subjects per session were randomly selected and paid for one of their decisions in the first part. ${ }^{19}$ The random selection took place at the end of the experiment to avoid income effects for the utilization behavior part of the experiment. In the utilization part, every subject was paid out, and the average payoff was 26.60 EUR.

The earnings were determined by the accumulated income from the utilization decision part of the experiment, which was the sum of periodic income after accounting for expenditures in sickness. Finally, subjects were asked to answer a short questionnaire with questions on demographics and questions related to their behavior in the previous decisions.

## 3. Results

### 3.1 Initial Aggregate Health Care Utilization Behavior

For the analysis of our results we focus predominantly on initial utilization behavior, the decisions to seek treatment or not, during the first 26 periods. First, the ceteris paribus assumption only holds over all conditions for this span since HighPricePeriods has only 26 periods. Second, we want to keep the spot price constant across conditions, and this is only the case before hitting the deductible: afterwards, the spot price would be 0 . Given our random draw of health events, the earliest possible period to hit the deductible of 600 is

[^8]period 30 if all 12 sicknesses are treated at this point. By period 26, 10 sickness events will have occurred, 4 sickness A events and 6 sickness B events (see Table 2).

## Low Price vs. High Price

We begin by comparing the results of LowPrice and HighPricePeriods. This resembles the empirical strategy utilized by Aron-Dine et al. (2015) to compare individuals who join the same deductible plan at different times of the year. By reducing the number of periods from 52 to 26 , while keeping everything else constant, subjects have less time to hit the deductible and face a higher future price. If subjects did not react to the future price, only the severe sickness A would be treated in each of the 4 periods it occurs within the first 26 periods, and we would not see a difference between the two conditions. If subjects behaved in a forwardlooking manner, we would expect treatment for all health events in LowPrice, resulting in 10 treatment decisions by period 26 .

Table 4.1 provides information on the average number of treatment decisions, regardless of the severity of illness, and the respective treatment rates for the severe sickness A and the mild sickness B for each condition. We see that the average number of treatment decisions by a subject is 8.15 (out of a possible 10) for LowPrice, while it is only 5.67 for HighPricePeriods over the first 26 periods. This difference is significant ( $p<0.001$ ) according to a two-sided Mann Whitney U-test (MWU). Thus, subjects decide to treat significantly more when the future price is low. We can also infer that the difference stems from treatment decisions for the mild health events, sickness B. Although over 95 percent of severe sickness A cases are treated in both conditions, indicating that subjects recognize this as a dominant strategy, treatment rates for sickness B differ substantially across the two conditions. In particular, 71 percent of the mild cases B are treated when the future price is low, even though not treating would be cheaper in a one shot situation ( 30 ECU vs. 50 ECU ). In contrast, only 31 percent are treated when the future price is higher, and the likelihood of spending beyond the deductible is low.

Table 4.1: Initial Behavior by Period 26 After 10 Sickness Cases

|  | Average Number <br> of Treatment <br> Choices | Treatment Rate for <br> Sickness A | Treatment Rate for <br> Sickness B |
| :--- | :---: | :---: | :---: |
| Part A: Main Conditions | 8.15 | 0.97 | 0.71 |
| LowPrice | 5.67 | 0.96 | 0.31 |
| HighPricePeriods | 5.04 | 0.94 | 0.22 |
| HighPriceDed |  |  |  |
| Part B: Robustness Conditions | 8.29 | 0.98 | 0.73 |
| LowPriceNoInfo | 5.45 | 0.92 | 0.3 |
| HighPriceDNoInfo | 8.1 | 0.9 | 0.75 |
| LowPriceReverse | 8.34 | 0.96 | 0.75 |
| LowPriceNeutral |  |  |  |

Notes. This table shows initial treatment behavior for sickness A and sickness B as well as the average number of treatment choices by period 26. Treatment rate indicates share of respective sickness cases treated by all subjects. See Table A.6.1 in Appendix A.6. for results after 52 periods.

Figure 2 illustrates this pattern over time. The green and orange bars reflect the treatments rates in LowPrice and HighPricePeriods by sickness periods, while the blue and the red line visualize the average respective future prices accounting for utilization behavior. It is apparent that subjects treat mild cases B less when the future price is high. Moreover, we observe a significant negative time trend for treating sickness B in HighPricePeriods. ${ }^{20} \mathrm{~A}$ higher future price due to a higher deductible in HighPriceDed also leads to a significantly lower number of treatments of 5.04 compared to LowPrice (MWU, $\mathrm{p}<0.0001$ ). This result indicates that subjects do not only react to the spot price but also show forward-looking behavior and anticipate that they will spend beyond the deductible in LowPrice.

[^9]Figure 2: Initial Utilization in LowPrice and HighPricePeriods


Result 1 corresponding to Hypothesis 2: A higher future price due to a lower number of periods in HighPricePeriods leads to a significantly lower number of treatments compared to LowPrice. The difference stems from different treatment rates for the mild sickness B. This indicates that subjects do not only react to the spot price but also show forward-looking behavior and anticipate that they will spend beyond the deductible in LowPrice.

Table 4.2 provides information on the average number of treatment choices by gender, as well as the expected number of treatment choices for a forward-looking or myopic individual for each condition. It shows that in all our LowPrice robustness conditions women treat sicknesses significantly less than men ( $p \leq 0.07$ ). This result is in line with Hayen et al. (2021) who show that women react stronger to cost-sharing schemes than men. Our results further suggest that women are more myopic than men.

Table 4.2: Initial Behavior by Period 26 After 10 Sickness Cases for Males and Females

|  | Average number of treatment choices |  |  | Expected number of treatment choices |  | $p$-value <br> (Males vs. <br> Females) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | All | Males | Females | Forwardlooking | Myopic |  |
| Part A: Main Conditions |  |  |  |  |  |  |
| LowPrice | 8.15 | 8.77 | 7.61 | 10 | 4 | 0.14 |
| HighPricePeriods | 5.67 | 4.44 | 6.4 | 4 | 4 | 0.05 |
| HighPriceDed | 5.04 | 5.24 | 4.57 | 4 | 4 | 0.22 |
| Part B: Robustness Conditions |  |  |  |  |  |  |
| LowPriceNoInfo | 8.29 | 8.87 | 7.76 | 10 | 4 | 0.07 |
| HighPriceDNoInfo | 5.45 | 6.5 | 4.71 | 4 | 4 | 0.27 |
| LowPriceReverse | 8.1 | 9.25 | 6.38 | 10 | 4 | 0.02 |
| LowPriceNeutral | 8.34 | 9.17 | 7.48 | 10 | 4 | 0.02 |

Notes. The table reports the average number of treatment choices for all subjects, and separately for males and females, by experimental condition. The reported exact $p$-values report differences between the average number of treatment choices are based on two-sided Mann-Whitney $U$ tests. We also present the expected number of treatment choices for risk-neutral forward-looking and risk-neutral myopic types.

## High Deductible vs. Fewer Periods

Do subjects react differently when manipulating the future price through either more decision periods or a higher deductible? We compare treatment behavior in HighPricePeriods with the behavior in HighPriceDed. If subjects reacted to the probability of hitting the deductible, the differences between both conditions should be marginal since the future price is almost identical. However, in HighPricePeriods the higher future price, or lower probability of hitting the deductible, is a product of halving the time to utilize the deductible, while in HighPriceDed the price was manipulated by increasing the deductible from 600 to 1150. We find that the average number of treatments is 5.67 in HighPricePeriods and 5.04 in HighPriceDed and the difference is not significant ( $p=0.38$ ). From Table 4.1 we again see that for both conditions the treatment rate in sickness A is close to 100 percent. The treatment rate for sickness B is similar between treatments: 31 percent in HighPricePeriods and 22 percent in HighPriceDed.

Figure 3: Initial Utilization in HighPricePeriods and HighPriceDed


Figure 3 illustrates the dynamic relationship between treatment decisions and future price in these two conditions. The blue and red lines in Figure 3 mark the average future prices for the two conditions over time, accounting for past utilization behavior in every period. Since the future price is almost identical initially, the lines effectively have an identical starting point. Over time, they stay relatively close together while approaching the price of 1, the point where it is impossible to hit the deductible, at period 26. The green and orange bars in Figure 3 represent the respective treatment rates in HighPricePeriods and HighPriceDed. The difference in behavior between the two possible health events is clearly visible, and for both HighPrice conditions the majority opts against treating when the mild sickness B occurs. This result further supports the conclusion that subjects react to changes in the future price. Also, subjects seem to display a good understanding of the future price in both HighPrice treatments. Whether the likelihood of hitting the deductible is manipulated via number of periods or level of deductible appears to be secondary. ${ }^{21}$

## Robustness Conditions: Information, Order and Framing Effects

To investigate the role of information on utilization behavior, we compare LowPrice and LowPriceNoInfo. In LowPrice subjects received information on their income, their cumulated

[^10]treatment costs, and their remaining deductible after each period. In LowPriceNoInfo the information was reduced and the cumulated income update was the only feedback that subjects received. When we compare the initial utilization between both conditions, the information about the remaining deductible and cumulated treatment costs does not seem to lead to differences in utilization behavior (see Table 4.1). Utilization is almost identical: 8.15 in LowPrice compared to 8.29 in LowPriceNoInfo, and the difference is not significant ( $p=0.86$ ). As with the average treatment rate, we find neither differences between the two treatment conditions nor a systematic pattern when considering the treatments rates per decision. Similarly, not giving information on the remaining deductible does not significantly affect the utilization when we compare HighPriceDed with HighPriceDNoInfo ( $p=0.70$ ).

Finally, we control for the order and framing. Reversing the order of part 1 and part 2 and eliciting preferences after the insurance part, does not have a significant effect. Utilization is 8.10 in LowPriceReverse and not significantly different from the benchmark condition ( $p=0.89$ ). In LowPriceNeutral we changed the wording of the instructions and removed any association to health, and framed the task as general indemnity insurance. The average utilization is 8.34 and not significantly different from LowPrice ( $p=0.71$ ).

### 3.2 Heterogeneity in Utilization Behavior

Our controlled experimental design allows for classifying individuals based on their observed individual utilization behavior. However, this is only possible in the conditions with a low future price since, assuming risk neutrality, the behavioral prediction is identical for both types given a high future price, not to treat sickness B. In the LowPrice conditions, however, decisions during sickness B provide a way to classify individuals. ${ }^{22}$ To begin with, we classify individuals assuming risk-neutrality and expected costs. In the following, we will then account for risk preferences and derive individually optimal treatment paths and strategies.

## Individual Utilization Behavior and Risk Neutrality

Assuming risk-neutrality and expected costs, only forward-looking individuals would treat the mild cases sickness B before hitting the deductible. From Table 5 we can infer that about 15 percent of subjects in the LowPrice conditions never treat the mild cases B and that

[^11]around 60 percent always treat sickness B. ${ }^{23}$ Hence, in a scenario with a low future price, forward-looking behavior appears to be prevalent.

Result 2 corresponding to Hypothesis 1: We find individuals who consistently never treat and individuals who consistently always treat sickness B in our LowPrice conditions. Hence, our results are consistent with some subjects looking towards the future price and taking that into account, and some subjects not only looking at the spot price.

Table 5: Classification of Treatment Behavior in LowPrice Conditions by Period 26

|  | Never Treat B | Mixed Strategy | Always treat B | Total |
| :--- | :---: | :---: | :---: | :---: |
| LowPrice <br> (pooled) | $24(14.7 \%)$ | $43(26.4 \%)$ | $96(58.9 \%)$ | 163 |

Notes. This table shows the classification of treatment behavior, indicated by the number of individuals, in the LowPrice conditions by period 26. The respective percentages are provided in parentheses.

However, recall that since this analysis does not account for individual risk preferences and thus we cannot disentangle whether an individual treats sickness B because the individual is forward-looking or because the individual is risk averse. Moreover, an optimal treatment path for an individual within a period depends on the previous events that have occurred as well as the individual's previous treatment decisions. For instance, for an individual who has previously deviated from their originally optimal path by not treating B in several sickness periods, it may at some later point become optimal not to treat sickness B anymore. We now derive such individually optimal strategies that account for individual risk preferences as well previous treatment behavior. ${ }^{24}$

## Individual Utilization Behavior and Risk Preferences

We estimate risk preferences at the individual subject level, to be able to make normative inferences that respect the heterogeneity of risk preferences that we expect a priori for individuals. Following Gao et al. (2023), we adopt a Bayesian approach to this estimation task, specifically a Bayesian Hierarchical Model (BHM). Our subjects made 20 binary

[^12]choices over lotteries with objective risks, and that number of observations would not be able to reliably generate individual estimates of risk preferences using the classical Maximum Likelihood methods used by Harrison and Ng (2016) for their normative evaluation of insurance choices. A BHM addresses this issue ${ }^{25}$ by pooling the behavior over all subjects to estimate hyper-parameters for a model of the risk preferences of a single "representative agent" that can then be used as informative priors for the estimation of risk preferences at the individual level. In this manner the larger data set of the subjects facing the same binary choices can be used to generate more plausible estimates for one individual than if each subject was estimated in isolation. The informative prior employed here is also referred to as a "shrinkage prior" since it effectively shrinks extreme estimates towards the pooled risk estimates. The extent of the shrinkage towards the pooled estimates depends on how informative the 20 observations are for each individual. For some individuals the prior will have little effect on the estimates, since their 20 observations are relatively informative about their risk preferences. But for other individuals, with more noisy behavior, the informative prior will play a more important role. In this manner the BHM is said to naturally "regularize" the estimates for each individual.

We estimate risk preferences for each individual assuming Expected Utility Theory (EUT) or the Rank Dependent Utility (RDU) model of Quiggin (1982). ${ }^{26}$ The latter has become one of the most important empirical generalizations of EUT. While the specification of the utility function is the same for EUT and RDU, RDU assumes that people may not necessarily behave as if given probabilities affect their lottery evaluations with objective values. Instead, they may distort these probabilities using a process that can be described by attaching subjective weights to probabilities. The RDU model derives probability weights from the entire distribution over ranked outcomes, not from individual probabilities, and avoids any theoretical violations of first-order stochastic dominance. The resulting decision weights reflect subjective distortions of objective probabilities. The RDU model requires the introduction of a probability weighting function. A variety of weighting functions have been

[^13]proposed in the literature. Prelec (1998) contributes a flexible two parameter specification of probability weighting: $\omega(p)=\exp \left[-\eta(-\ln p)^{\phi}\right]$, with $\eta>0$ and $\phi>0$. This weighting function is derived from several axioms that reflect apparent regularities of probability weighting, and requires the estimation of two additional parameters $\eta$ and $\phi .{ }^{27}$

Since EUT is nested within RDU when $\eta=\phi=0$, we could just estimate risk preferences assuming RDU. But for many policy-makers and economists, EUT is more normatively attractive than RDU for policy evaluation. We remain agnostic on that issue, for reasons explained in Harrison and Ross (2018; p.49ff.), but prefer here to be able to evaluate welfare using EUT and RDU to be able to inform policy with either, and to see if the use of either model makes much difference to normative conclusions. For both EUT and RDU we assume a Constant Relative Risk Aversion utility function, and for RDU we assume a flexible two-parameter probability weighting function due to Prelec (1998). We evaluate behavior using the mean of the posterior distributions of each individual for their risk preferences parameters.

In addition, to allow for the possibility that some individuals might make some choices using EUT risk preferences and other choices using RDU risk preferences, we consider the effect of allowing each individual to use risk preferences from a mixture of EUT and RDU preferences. Following Harrison and Rutström (2009), each individual has some probability of each choice being based on EUT or RDU preferences, and we estimate this probability. Over all subjects the average probability of using EUT risk preferences is only 0.23 , but individual estimates for this parameter vary between 0.01 and 0.98 .

## Evaluation of Treatment Choices

In Section 3.2. we classified utilization behavior based on the observed treatment choices in our LowPrice conditions assuming risk-neutrality and expected costs. Having characterized subjects by their elicited risk preferences, we can now use their individual risk parameters to derive optimal treatment choices. This allows us to reconcile actual choices made with these optimal choices. We can then identify myopic behavior or decision errors that cannot be explained by individual risk preferences and thus also forward-looking behavior.

For evaluating the series of treatment choices, we calculate the prospective expected costs and expected utilities according to the mixture of EUT and RDU or EUT for each individual in each respective period for different treatment strategies. We evaluate the two strategies (i)

[^14]either always treat both sicknesses A and B or (ii) only treat sickness $\mathrm{A}^{28}$ for the remaining duration as potential optimal strategies. For each period and individual, we calculate the prospective expected utility under the mixture of EUT and RDU and EUT if the individual would stick to one of these two strategies until the end. While in period 0 , i.e. before making any treatment choices, a subject facing a deductible of 600 knows for sure that he will have to pay out of pocket for the first 12 sicknesses. The probability to pay out of pocket in later stages of the experiment depends on the previous treatment strategy of the subject as this affects the probability to hit the deductible. We use these probabilities to hit the deductible as well as the treatment cost of 50 or the opportunity cost of 30 , to calculate the prospective expected costs and individual utilities for each prospective period and sum it up. ${ }^{29}$ This gives us the cumulated expected utility of the two strategies for each subject in period 0 . We then repeat this exercise in every period, taking the actual utilization up to this point into account. This leave us with the expected utilities for both strategies for every subject in every period of the game which allows us to evaluate the optimal treatment choices and thus an optimal treatment path for each subject.

Given actual treatment choices, we can then determine whether an individual deviates from the optimal choice in the respective period. We assume that individuals deviate from their mixture of EUT and RDU or EUT optimal path in case they don't treat sickness A in the respective periods, or do not treat sickness $B$ although this would be optimal for them in the respective period, or do treat sickness $B$ although this would not be optimal for them in the respective period. If someone is an optimal mixture of EUT and RDU or EUT decision maker it hence does not follow that they never or always treat B in every period. In fact, the optimal path for an individual might be a mixed strategy. The total number of these deviations serves as a measure of decision error that we can relate to our previous classification in the LowPrice conditions that was based on whether the mild sickness B was treated in the initial 26 periods. When a deviation from an optimal choice and the decision not to treat the mild sickness B overlap, it is very likely due to a lack in forward-looking behavior and not due to risk preferences. If the choice not to treat the mild sickness B, however, is in line with the

[^15]preferences (no deviation from the optimal choice), we cannot rule out that the individual considered the future price and thus cannot label it as myopic.

For this, we now consider the first 26 periods and the mild sicknesses B. Given our exante random draw, every subject faced 6 cases of sickness B. From Table 6 , we can see the number of individuals for whom the previous classification based on risk neutrality and expected costs resembles their optimal decision based on their risk preferences in each category as well as the average number of deviations from the optimal choice. Among the subjects who never treat the mild sickness $\mathrm{B}, 17$ (8) out of 24 individuals do so in line with their preferences under the mixture of EUT and RDU (EUT). For them not treating sickness B is not a behavioral bias or lack in forward-looking behavior. On the other hand, 50 (88) out of 96 subjects behave optimally by treating both sickness cases under the mixture of EUT and RDU (EUT). Inconsistent play, i.e. switching between treating and not treating B, is optimal for 5 (17) of the 43 subjects under the mixture of EUT and RDU (EUT) model. From the average number of deviations, we can also see that those who never treat sickness B are much less (more) prone to deviate and make behavioral mistakes under the mixture of EUT and RDU (EUT). Their average number of deviations is 1.46 (2.71) while for those who always treat sickness B it is 2.46 ( 0.21 ) under the mixture of EUT and RDU (EUT). (See A. 5 Figures A.5.1 and A.5.2 for histograms with distributions for the mixture of EUT and RDU and EUT). Assuming a mixture model of EUT and RDU (EUT) risk preferences, a classification of observed treatment behavior based on risk neutrality seems to work quite well for the strategy to never treat sickness B (to always treat sickness B) but is suboptimal for the strategy to always treat sickness B (never treat sickness B).

Table 6 Number of Optimal Decision Makers and Average Number of Deviations at Period 26 (LowPrice Conditions Pooled)

|  | Never Treat B | Mixed Strategy | Always Treat B |
| :--- | :---: | :---: | :---: |
| Optimal Decision Makers Mixture | 17 out of 24 | 5 out of 43 | 50 out of 96 |
| Avg. \# of deviations Mixture | $1.46(\operatorname{sd} 2.4)$ | $2.51(\operatorname{sd} 1.62)$ | $2.46(\operatorname{sd} 2.79)$ |
| Optimal Decision Makers EUT | 8 out of 24 | 17 out of 43 | 88 out of 96 |
| Avg. \# of deviations EUT | $2.71(\operatorname{sd~2.48)~}$ | $1.74(\operatorname{sd~1.72)}$ | $0.21(\operatorname{sd} 0.75)$ |

Notes. This table shows the number of optimal decision makers and the average number of deviations when accounting for individual risk preferences based on a mixture model of EUT and RDU and an EUT model. Standard deviations (sd) are reported in parentheses.

Accounting for risk preferences, the optimal treatment paths now allow for classifying individuals by their degree of forward-looking behavior. For this, we assume that individuals that always deviate from their optimal path in periods with sickness B are fully myopic, while those who never deviate from their optimal path in these periods are fully forward-looking. Those individuals who deviate from their optimal path several times but not always are mixed types in the sense that they display some degree of forward-looking behavior. Table 7 displays the number (percentages) of the respective behavioral types in our LowPrice conditions. These results show that under the mixture model of EUT and RDU risk preferences we find a more equal distribution of myopic ( $23 \%$ ), mixed ( $33 \%$ ) and forwardlooking types (44\%) than under the EUT model of risk preferences with about two thirds of forward-looking types, one third of mixed types and only very few myopic types. These results suggest that the distribution of types differs substantially depending on which model of risk preferences one assumes to be the normative metric. One reason for the difference in the distribution of behavioral types between the mixture model of EUT and RDU risk preferences and EUT preferences is that RDU allows distortions of probabilities compared to EUT, and this can be expected to have a "first-order" effect on the assignment to behavioral types since the types are defined by how the individual evaluates the probabilities of choices on future options.

Table 7 Classification of Behavioral Type by Number of Deviations at Period 26 and Gender (LowPrice Conditions Pooled)

|  | Myopic | Mixed | Forward-looking | Total |
| :---: | :---: | :---: | :---: | :---: |
| All |  |  |  |  |
| Total Mixture | $\begin{gathered} 37 \\ (23 \%) \end{gathered}$ | $\begin{gathered} 54 \\ (33 \%) \end{gathered}$ | $\begin{gathered} 72 \\ (44 \%) \end{gathered}$ | 163 |
| Total EUT | $\begin{gathered} 5 \\ (3 \%) \\ \hline \end{gathered}$ | $\begin{gathered} 45 \\ (28 \%) \end{gathered}$ | $\begin{gathered} 113 \\ (69 \%) \end{gathered}$ | 163 |
| Male |  |  |  |  |
| Total Mixture | $\begin{gathered} 20 \\ (25 \%) \end{gathered}$ | $\begin{gathered} 19 \\ (23 \%) \end{gathered}$ | $\begin{gathered} 42 \\ (52 \%) \end{gathered}$ | 81 |
| Total EUT | $\begin{gathered} 1 \\ (1 \%) \end{gathered}$ | $\begin{gathered} 14 \\ (17 \%) \end{gathered}$ | $\begin{gathered} 66 \\ (82 \%) \end{gathered}$ | 81 |
| Female |  |  |  |  |
| Total Mixture | $\begin{gathered} 17 \\ (21 \%) \end{gathered}$ | $\begin{gathered} 35 \\ (43 \%) \end{gathered}$ | $\begin{gathered} 30 \\ (36 \%) \end{gathered}$ | 82 |
| Total EUT | $\begin{gathered} 4 \\ (5 \%) \\ \hline \end{gathered}$ | $\begin{gathered} 31 \\ (38 \%) \\ \hline \end{gathered}$ | $\begin{gathered} 47 \\ (57 \%) \\ \hline \end{gathered}$ | 82 |

Notes. This table shows the number of participants for each behavioral type by the number of deviations for the pooled LowPrice condtions at period 26 for all subjects, by gender, and by a mixture of EUT and RDU and EUT. The percentage share of the respective behavioral types is provided in parentheses.

Table 7 also shows that the distribution of types differs by gender. Under the mixture model of EUT and RDU risk preferences we find $25 \%$ ( $21 \%$ ) myopic, $23 \%$ ( $43 \%$ ) mixed and $52 \%(36 \%)$ forward-looking types for males (females). Under an EUT model of risk preferences we find $1 \%$ (5\%) myopic, $17 \%$ ( $38 \%$ ) mixed and $82 \%$ ( $57 \%$ ) forward-looking types for males (females). Hence, under both a mixture model of EUT and RDU risk preferences and one of EUT preferences there is a tendency towards more male (female) forward-looking (mixed) types.

Result 3: Accounting for individual risk preferences, we identify consistent myopic and forward-looking behavior. The distribution of types differs depending on which model of risk preferences one assumes to be the normative metric: there are substantially more (less) forward-looking (myopic) types under the EUT model of risk preferences than under the mixture model of EUT and RDU risk preferences.

We further investigate determinants for not treating according to an individual mixture of EUT and RDU or EUT preferences. We run a random effect probit to describe the determinants of deviating from the optimal choice over the first 26 periods in all LowPrice conditions, see Table 8. Under a mixture of EUT and RDU we find significant effects on the deviation from the optimal choice resulting from the experimental condition of giving individuals information on the remaining deductible level after each round. The latter significantly reduces the likelihood to deviate from the optimal path. In contrast, under EUT we find a strong gender effect: women have a higher likelihood for deviating from their optimal choice. Hence women appear to make more behavioral mistakes in the sense that they do not act in accordance to their risk preferences under EUT.

Table 8 Deviations from Optimal Paths and Characteristics at period 26 (LowPrice Conditions Pooled)

|  | Mixture |  | EUT |  |
| :--- | :---: | :---: | :---: | :---: |
|  | Deviate | $p$-value | Deviate | $p$-value |
| female | 0.159 | 0.217 | 0.74 | 0.001 |
| info | -0.503 | 0.001 | 0.193 | 0.425 |
| Priv. healthinsurance | -0.086 | 0.616 | 0.324 | 0.202 |
| neutral | 0.110 | 0.486 | 0.029 | 0.9 |
| $N$ | 4238 | 4238 | 4238 | 4238 |
| $N$ |  |  |  |  |

Notes. This table provides the results of a random effect probit regression used to describe the determinants of deviating from the optimal choice in LowPrice conditions for mixture and EUT with Deviate being the regression coefficient. Based on 26 decisions made by 163 subjects. $p$-values are reported in adjacent columns.

## Welfare effects

In a next step we undertake an evaluation of utilization choices in terms of their welfare effects, using the notion of the Expected Consumer Surplus (ECS) of observed utilization choice. The logic is conventional in terms of welfare economics, with one extension. Each utilization choice offers the subject a well-defined lottery. How the subject trades off the final outcomes of that lottery is determined by her utility function parameters, and how the subject weights the utility of different payoffs is determined by the probabilities of those payoffs and her probability weighting parameters. Assume for the moment that she is an EUT decisionmaker, solely to ease the basic exposition. Then if we know her utility function parameters we can infer the EUT that she attaches to each possible choice. We can also then immediately calculate the Certainty Equivalent ( $C E$ ) of that lottery to her, by solving for $C E$ in the equation $U(C E)=$ EUT. To take simple case, if we had used the power utility function $U(x)=x^{r}$, then the CE is equal to $E U T^{(1 / r)}$. Even if closed-form expressions for the $C E$ do not exist, it is a simple numerical matter to find the scalar $C E$ that solves this equation. If the decision maker chooses one utilization lottery A over another utilization lottery B, we just evaluate $E U T^{A}$ and $E U T^{B}$, then evaluate $C E^{A}$ and $C E^{B}$, and the ECS is simply $C E^{A}-C E^{B}$. In the familiar language from welfare economics, the ECS is just the certain amount of money that the individual would require to be just willing to give up her preferred utilization choice for the alternative.

All of this is familiar welfare economics, assuming we know the risk preferences of the individual or, as in our case, we can predict those risk preferences from a pooled model with demographic characteristics that differentiate each individual. We can extend it immediately to the case of the mixture model of EUT and RDU, where each individual has a distinct implied coefficient for a respective model parameter. The welfare-theoretic logic is identical and standard.

Some economists insist that only EUT risk preferences are normatively attractive, and it is a simple matter to substitute EUT parameters for that individual. We disagree with this assumption about EUT being normatively attractive, for reasons discussed in Harrison and Ng (2016), but that is a debate for another time, and for now we can consider the effects on our conclusions of also assuming that EUT risk preferences are the appropriate normative metric.

However, what is novel here is that we have a measure of the risk preferences of the individual that is independent of the observed utilization choice. For normative evaluation of the utilization choice, we must in fact have some independent measure of risk preferences. The reason is that if we inferred risk preferences from observed utilization choices, we would
always infer, in expectation, that the ECS from the observed utilization choice was zero or positive by direct revealed preference. In the example above, we were careful to say that we observe the subject choosing utilization lottery A over B. We did not say that the ECS of that decision was positive. In fact, and this is the normative point of behavioral welfare evaluation, we might have observed the subject making a poor decision and generating a negative ECS. ${ }^{30}$ This approach to the normative evaluation of lotteries was developed by Harrison and Ng (2016), and reviewed methodologically by Harrison (2019).

Once we have evaluated the ECS for each individual and choice, we can collate these effects over all of the choices made by an individual. This measures Efficiency, in the language of experimental economists since Plott and Smith (1978): how much of the potential ECS did the subject actually make from her observed choices. The usual measures of Efficiency were developed for non-stochastic settings, in terms of the ability of the subject to "extract surplus" from the experimenter by appropriate choices. We simply utilize ECS and ask about the ability of the subject to extract surplus in expectation.

Figure 4 displays results for ECS in the left panel, and results for Efficiency in the right panel. We present results assuming a mixture model of EUT and RDU risk preferences in red, and assuming EUT risk preferences in blue. ${ }^{31}$ Under both the mixture of EUT and RDU and EUT we see a large number of small welfare losses and a long tail of larger losses. The number of the latter is slightly higher for EUT than for the mixture of EUT and RDU along the tail. These findings for ECS translate into a single mode for Efficiency under the mixture of EUT and RDU at high levels around $80 \%$ to $100 \%$, and a slight tail of lower efficiency levels. For EUT we observe significant number of subjects around 50 to $70 \%$ Efficiency, and then some subjects with between 0 and $25 \%$ Efficiency. These results point to more individuals making mistakes that were welfare costly under EUT, to the point where their Efficiency drops well below $50 \%$ in many cases. Under EUT we find that more than two thirds of the subjects are classified forward-looking. Hence, these costly mistakes appear to originate from a few subjects making very large mistakes.

[^16]Figure 4: Total Foregone Welfare


> - EUT - - . Mixture of EUT \& RDU

We can explore these welfare results by examining the "marginal effects" of treatments or demographics on welfare distributions. In this case we consider the task and subject characteristics shown in Figure 5 for the mixture of EUT and RDU. The diamond symbol is the point estimate of the marginal effect, and the bars either side of that symbol show the $95 \%$ confidence interval.

For the mixture of EUT and RDU and EUT we find a significant effect of the low probability of reaching a deductible leading to a substantial welfare loss. The loss is stronger under EUT, and appears to be the driving factor behind larger numbers in the tail of losses in ECS noted earlier under EUT for Figure 4. We also identify welfare effects from the variations in our robustness conditions. Under the mixture of EUT and RDU we find that there is a tendency for extra information on the remaining level of the deductible to lower welfare. Under EUT we find only a tiny, negative effect of extra information on welfare. For the abstract insurance context, we find effects that go into different directions. Given a mixture of EUT and RDU we find that an abstract insurance framing leads to higher welfare, while given EUT it has a slightly lower welfare.

Figure 5: Marginal Effects on Efficiency under a Mixture of EUT and RDU
Marginal average effects from a Fractional regression model


Task Characteristics
-Low probability of reaching deductible
Extra information
Abstract context

## Subject Characteristics

-Female
Age in years (over 17)
Field of study is Economics
Has private health insurance
Has own statutory health plan

Percentage Point Effect on Efficiency

Figure 6: Marginal Effects on Efficiency under EUT
Marginal average effects from a Fractional regression model


Task Characteristics
Low probability of reaching deductible
Extra information
Abstract context

## Subject Characteristics

Female
Age in years (over 17)
Field of study is Economics
-Has private health insurance
Has own statutory health plan

Figures 5 and 6 also show that there are welfare effects which depend on individual subject characteristics. Under the mixture of EUT and RDU we find a significant welfare reduction for women. In contrast, under EUT there is a significant increase in welfare for women. Studying in the field of economics, having a private health insurance, and having
statutory health insurance all have a negative effect on welfare under the mixture of EUT and RDU and a positive effect under EUT.

Result 4: There is a substantial welfare loss due to a lack in forward-looking behavior irrespective of whether we assume a mixture model of EUT and RDU risk preferences or EUT risk preferences. The distribution and drivers of the welfare losses differ for the two models of risk preferences.

## 4. Limitations

Our results have some limitations, relating to the general concern about the external validity of the results when using experimental data from controlled laboratory experiments.

First, one limitation might be that in our experimental design, the cost of health care and the probabilities of sickness occurrence are known to each individual. In contrast, in the field, both aspects are often unknown or uncertain. Although we consider this deviation from the field to be the strength of our experimental approach, since it allows the application of familiar economic theory, we cannot say anything about the relationship between uncertainty and forward-looking behavior. ${ }^{32}$

A further limitation could be the health framing of the experimental setting. By introducing a control condition with a neutral framing, we checked whether there were any differences in treatment behavior. Although one might expect the health framing to lead to higher treatment rates, as some individuals place a relatively high non-monetary value on health, we did not find any differences. However, this could be due to the health framing not being strong enough: the wording was provided in a health frame where the outcomes were monetary and unrelated to the personal health of the individuals. In the field a strong nonmonetary value for health might lead to larger treatment rates and more forward-looking types.

Third, another limitation relates to the way we designed the parameterization, information and understanding of the dynamic incentives within our experimental setting. Our design followed the empirical strategy of Aron-Dine et al. (2015) by varying the expected end-ofyear price while holding the spot price constant. For this, we implemented the LowPrice and HighPrice conditions. We deliberately designed the parameterization in the LowPrice

[^17]conditions in such a way that individuals are confronted with strong dynamic incentives: they have a high probability of reaching the deductible. In contrast, in the HighPrice conditions they deliberately have a relatively low probability of reaching the deductible. This makes it easy to calculate the expected treatment costs and thus the probability of reaching the deductible threshold in the LowPrice conditions, but rather difficult in the HighPrice conditions. The differences in the HighPrice conditions could therefore also be attributed to cognitive abilities in the sense of abilities to calculate the deductible threshold accurately. However, we checked the calculation skills of individuals as far as possible: all participants were asked control questions ex ante the actual decisions. We included two control questions on the understanding of the deductible: one which asks about the level of the deductible and one that asks about the costs one has to bear if the deductible is reached. ${ }^{33} \mathrm{We}$ also included a question about the opportunity cost of not treating a sickness. Moreover, participants received information after each period about their income and the remaining amount of the deductible in all conditions except the control condition without information. An interesting avenue for future research could be to follow the empirical approach of Guo and Zhang (2019) and test for fully forward-looking behavior by varying the spot price while fixing the expected end-ofyear price. This could possibly make the design simpler.

A final limitation relates to the fact that we base our welfare evaluation on an exogenous measure of risk preferences that was elicited in a neutral frame. While this approach has the clear advantage of avoiding problems associated with inferring risk preferences from observed decisions, it also comes with its own challenges. One challenge could be context dependency, i.e. the fact that individuals make different decisions in different contexts. In our setting, this context dependency could even be amplified, since we used a health framing for the choice task and a neutral framing for the risk preference task. Further research is needed to better understand the relationship between risk preferences, their framing, and the utilization of healthcare consumption.

## 5. Conclusions

We complement the empirical evidence on the effects of nonlinear deductible contracts on health care utilization behavior by using a controlled laboratory experiment. Compared to the field, the laboratory environment allowed us to control for keeping the spot price constant

[^18]while varying the future price, as well as for other confounding factors such as seasonality, liquidity, or comorbidities. Subjects in the experiment went through a cycle of periods and were insured by a health plan with a deductible. In each period, they faced probabilistic health events and had to choose between treating or not. We also elicited individual risk preferences for each subject, allowing for welfare estimation of observed health care utilization behavior.

Consistent with recent empirical results for nonlinear health insurance plans with deductibles, and Medicare Part D plans, we find that subjects respond to the embedded dynamic incentives in aggregate and do not only react to the spot price. We also find a tendency for women to treat significantly less sicknesses under a deductible plan with a low future price. This is consistent with Hayen et al. (2021) who show that women react more strongly to cost-sharing than men. Whether the future price is manipulated through more decision periods or a higher deductible does not significantly affect utilization behavior as long as the likelihood of hitting the deductible is the same.

Aron-Dine et al. (2015) suggest that people understand the dynamic incentives of health insurance contracts with deductibles to some degree. This implies that both the spot price of insurance as well as the future price of insurance should be relevant to determine the price elasticities of demand for medical services. We contribute to the characterization of this heterogeneity. Specifically, our results suggest that, assuming a mixture of EUT and RDU (EUT) model of risk preferences, $23 \%$ ( $3 \%$ ) of individuals are myopic and only take the spot price into account, $33 \%(28 \%)$ are mixed in their choice behavior and $44 \%(69 \%)$ are forward-looking and take the future price of insurance into account.

Our results further show that the drivers of welfare effects also crucially depend on the model of risk preferences one assumes to be the normative benchmark. The results from our treatment of giving individuals additional information on the remaining level of the deductible in the robustness conditions also provides some insights to policy makers. Assuming a mixture model of EUT and RDU risk preferences we show that regularly giving individuals information on the remaining level of the deductible may actually decrease efficiency, while assuming EUT we find only a tiny negative effect. Previous studies suggest that providing individuals with information, or simplifying the decision process, can indeed affect decision outcomes. We add to this literature and show that the extend of the effects of such policy measures can depend on the underlying model of risk preferences used for normative evaluation.

A further driver of welfare that depends on the underlying model of risk preferences is gender. Under a mixture model of EUT and RDU risk preferences, we find a significant
welfare reduction for women, whereas there is a significant welfare increase for women under an EUT model of risk preferences. We thereby add to the literature on gender differences in health care consumption given a deductible. The tendency we find for women to treat less sicknesses under a deductible, which is consistent with the observations in the field, may translate either into a positive or negative welfare impact depending on the underlying model of risk preferences used for normative evaluation.

Future research could investigate an extension of the horizon of the experiment and hence also account for individual time preferences, or consider subjective beliefs about loss probabilities.

## References

Abaluck, J., Gruber, J., \& Swanson, A. (2018). Prescription drug use under Medicare Part D: A linear model of nonlinear budget sets. Journal of Public Economics, 164, 106-138.

Andersen, S., Harrison, G. W., Lau, M. I., \& Rutström, E. E. (2006). Elicitation using multiple price list formats. Experimental Economics, 9(4), 383-405.

Andersen, S., Harrison, G. W., Lau, M. I., \& Rutström, E. E. (2008). Eliciting risk and time preferences. Econometrica, 76(3), 583-618.

Aron-Dine, A., Einav, L., \& Finkelstein, A. (2013). The RAND Health Insurance Experiment, three decades later. Journal of Economic Perspectives, 27(1),197-222.

Aron-Dine, A., Einav, L., Finkelstein, A., \& Cullen, M. R. (2012). Moral hazard in health insurance: How important is forward looking behavior? Working Paper 17802. National Bureau of Economic Research.

Aron-Dine, A., Einav, L., Finkelstein, A., \& Cullen, M. (2015). Moral hazard in health insurance: do dynamic incentives matter? Review of Economics and Statistics, 97(4), 725741.

Arrow, K. (1971). Essays in the Theory of Risk Bearing. Amsterdam: North-Holland Publishing Company.

Bhargava, S., \& Manoli, D. (2015). Psychological frictions and the incomplete take-up of social benefits: Evidence from an IRS field experiment. American Economic Review, 105(11), 3489-3529.

Bhargava, S., Loewenstein, G., \& Sydnor, J. (2017). Choose to lose: Health plan choices from a menu with dominated option. Quarterly Journal of Economics, 132(3), 1319-1372.

Biener, C., \& Zou, L. (2022). The Value of Choice in Mandatory Health Insurance. Working Papers on Finance 2021-04, University of St. Gallen, School of Finance.

Brot-Goldberg, Z. C., Chandra, A., Handel, B. R., \& Kolstad, J. T. (2017). What does a deductible do? The impact of cost-sharing on health care prices, quantities, and spending dynamics. Quarterly Journal of Economics, 132(3), 1261-1318.

Cardon, J. H. \& Hendel, I. (2001). Asymmetric information in health insurance: evidence from the National Medical Expenditure Survey. RAND Journal of Economics, 32(3), 408427.

Dalton, C. M., Gowrisankaran, G., \& Town, R. J. (2020). Salience, myopia, and complex dynamic incentives: Evidence from Medicare Part D. Review of Economic Studies, 87(2), 822-869.

Einav, L., Finkelstein, A., \& Schrimpf, P. (2015). The response of drug expenditure to nonlinear contract design: Evidence from Medicare Part D. Quarterly Journal of Economics, 130(2), 841-899.

Ellis, R. P. (1986). Rational behavior in the presence of coverage ceilings and deductibles. RAND Journal of Economics, 17(2): 158-175.

Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J.P., Allen, H., Baicker, K. \& Oregon Health Study Group (2012). The Oregon Health Insurance Experiment: evidence from the first year. Quarterly Journal of Economics, 127(3), 10571106.

Gao, X. S., Harrison, G. W., \& Tchernis, R. (2023). Behavioral welfare economics and risk preferences: A Bayesian approach. Experimental Economics, 26(2), 273-303.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. Journal of the Economic Science Association, 1(1), 114-125.

Guo, A., \& Zhang, J. (2019). What to expect when you are expecting: Are health care consumers forward-looking? Journal of Health Economics, 67, 102216.

Harrison, G. W. (2019). The behavioral welfare economics of insurance. The Geneva Risk and Insurance Review, 44(2), 137-175.

Harrison, G. W. (2024). Risk Preferences and Risk Perceptions in Insurance Experiments: Some Methodological Challenges. CEAR Working Paper 2024-0, Robinson College of Business, Georgia State University and also Geneva Risk and Insurance Review, forthcoming.

Harrison, G., Morsink, K., \& Schneider, M. (2022). Literacy and the quality of index insurance decisions. Geneva Risk and Insurance Review, 47(1), 66-97.

Harrison, G.W., \& Ng, J. M. (2016). Evaluating the expected welfare gain from insurance. Journal of Risk and Insurance, 83(1), 91-120.

Harrison, G. W., \& Ross, D. (2018). Varieties of paternalism and the heterogeneity of utility structures. Journal of Economic Methodology, 25(1), 42-67.

Harrison, G. W., \& Ross, D. (2023). Behavioral welfare economics and the quantitative intentional stance. Models Of Risk Preferences: Descriptive And Normative Challenges (767). Emerald Publishing Limited (Research in Experimental Economics, volume 23), Bingley UK.

Harrison, G. W. , \& Rutström, E. E. (2008). Risk aversion in the laboratory. In Cox, J.C. and Harrison, G.W., eds., Risk Aversion in Experiments (41-196). Emerald Publishing Limited (Research in Experimental Economics, volume 12), Bingley UK.

Harrison, G. W., \& Rutström, E. E. (2009). Expected utility theory and prospect theory: One wedding and a decent funeral. Experimental Economics, 12(2), 133-158.

Harrison, G. W., \& Swarthout, J. T. (2023). Cumulative prospect theory in the laboratory: A reconsideration. In Harrison, G.W. and Ross, D., eds., Models of Risk Preferences: Descriptive and Normative Challenges (107-192). Emerald Publishing Limited (Research in Experimental Economics, volume 23), Bingley UK.

Hastings, J. S., \& Weinstein, J. M. (2008). Information, school choice, and academic achievement: Evidence from two experiments. Quarterly Journal of Economics, 123(4), 1373-1414.

Hayen, A. P., Klein, T. J., \& Salm, M. (2021). Does the framing of patient cost-sharing incentives matter? The effects of deductibles vs. no-claim refunds. Journal of Health Economics, 80, 102520.

Hermanns, B., Kairies-Schwarz, N., Kokot, J., \& Vomhof, M. (2023). Heterogeneity in health insurance choice: an experimental investigation of consumer choice and feature preferences. SSRN Working Paper.

Holt, C. A. \& Laury, S. K. (2002). Risk aversion and incentive effects. American Economic Review, 92(5), 1644-1655.

Johansson, N., de New, S. C., Kunz, J. S., Petrie, D., \& Svensson, M. (2023). Reductions in out-of-pocket prices and forward-looking moral hazard in health care demand. Journal of Health Economics, 87, 102710.

Kairies-Schwarz, N., Kokot, J., Vomhof, M., \& Weßling, J. (2017). Health insurance choice and risk preferences under cumulative prospect theory - an experiment. Journal of Economic Behavior \& Organization, 137, 374-397.

Keeler, E. B., Newhouse, J. P., \& Phelps, C. E. (1977). Deductibles and the demand for medical care services: The theory of a consumer facing a variable price schedule under uncertainty. Econometrica, 45, 641-655.

Keeler, E. B. \& Rolph, J. E. (1988). The demand for episodes of treatment in the health insurance experiment. Journal of Health Economics, 7(4), 337-367.

Klein, T. J., Salm, M., \& Upadhyay, S. (2022). The response to dynamic incentives in insurance contracts with a deductible: Evidence from a differences-in-regressiondiscontinuities design. Journal of Public Economics, 210, 104660.

Kling, J. R., Mullainathan, S., Shafir, E., Vermeulen, L. C., \& Wrobel, M. V. (2012). Comparison friction: Experimental evidence from Medicare drug plans. Quarterly Journal of Economics, 127(1), 199-235.

Kowalski, A. E. (2015). Estimating the tradeoff between risk protection and moral hazard with a nonlinear budget set model of health insurance. International Journal of Industrial Organization, 43, 122-135.

Krieger, M., \& Felder, S. (2013). Can decision biases improve insurance outcomes? An experiment on status quo bias in health insurance choice. International Journal of Environmental Research and Public Health, 10(6), 2560-2577.

Levitt, S. D., \& List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world?. Journal of Economic Perspectives, 21(2), 153-174.

Loewenstein, G., Friedman, J.Y., McGill, B., Ahmad, S., Linck, S., Sinkula, S., Beshears, J., Choi, J.J., Kolstad, J., Laibson, D., Madrian, B.C., List, J.A., \& Volpp, K.G. (2013). Consumers' misunderstanding of health insurance. Journal of Health Economics, 32(5), 850-862.

Mimra, W., Nemitz, J., \& Waibel, C. (2020). Voluntary pooling of genetic risk: A health insurance experiment. Journal of Economic Behavior \& Organization, 180, 864-882.

Pauly, M. V. (1968). The economics of moral hazard: comment. The American Economic Review, 58(3), 531-537.

Plott, C. R., \& Smith, V. L. (1978). An experimental examination of two exchange institutions. The Review of Economic Studies, 45(1), 133-153.

Prelec, D. (1998). The Probability Weighting Function. Econometrica, 66(3), 497-528.
Quiggin, J. (1982). A Theory of Anticipated Utility. Journal of Economic Behavior \& Organization, 3(4), 323-343.

Samek, A., \& Sydnor, J. R. (2020). Impact of Consequence Information on Insurance Choice. Working Paper 28003. National Bureau of Economic Research.

Schram, A., \& Sonnemans, J. (2011). How individuals choose health insurance: An experimental analysis. European Economic Review, 55(6), 799-819.

Zeckhauser, R. (1970). Medical insurance: A case study of the tradeoff between risk spreading and appropriate incentives. Journal of Economic Theory, 2(1), 10-26.

## A Appendix

## A. 1 Experimental Design

## A.1.2 Price Lists for Risk Preferences

The following tables display the tasks subjects had to complete for the elicitation of risk preferences. Subjects had to decide between Option A and Option B in every row. The (expected) payoff differences were not displayed to them. Each subject chose one line at which to switch from A to B: This is called a sequential Multiple Price List (sMPL) by Andersen et al. (2006).

At the end of the experiment one person per session was randomly selected to get paid for one random decision in the risk task. This procedure happened at the very end of the experiment to avoid income effects that could confound behavior in the second part.

Table A.1.1: Risk Lottery A

|  | Option A |  | Option B |  | Expected payoff <br> difference |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | 20 EUR | 16 EUR | 38.50 EUR | 1 EUR |  |
| 1 | 10\% | 90\% | 10\% | 90\% | 11.65 EUR |
| 2 | 20\% | 80\% | 20\% | 80\% | 8.30 EUR |
| 3 | 30\% | 70\% | 30\% | 70\% | 4.95 EUR |
| 4 | 40\% | 60\% | 40\% | 60\% | 1.60 EUR |
| 5 | 50\% | 50\% | 50\% | 50\% | -1.75 EUR |
| 6 | 60\% | 40\% | 60\% | 40\% | -5.10 EUR |
| 7 | 70\% | 30\% | 70\% | 30\% | -8.45 EUR |
| 8 | 80\% | 20\% | 80\% | 20\% | -11.80 EUR |
| 9 | 90\% | 10\% | 90\% | 10\% | -15.15 EUR |
| 10 | 100\% | 0\% | 100\% | 0\% | -18.50 EUR |

Table A.1.2: Risk Lottery B

|  | Option A |  | Option B |  | Expected payoff |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | 22.50 EUR | 15 EUR | 40 EUR | 5 EUR | difference |
| 1 | 10\% | 90\% | 10\% | 90\% | 7.25 EUR |
| 2 | 20\% | 80\% | 20\% | 80\% | 4.50 EUR |
| 3 | 30\% | 70\% | 30\% | 70\% | 1.75 EUR |
| 4 | 40\% | 60\% | 40\% | 60\% | -1.00 EUR |
| 5 | 50\% | 50\% | 50\% | 50\% | -3.75 EUR |
| 6 | 60\% | 40\% | 60\% | 40\% | -6.50 EUR |
| 7 | 70\% | 30\% | 70\% | 30\% | -9.25 EUR |
| 8 | 80\% | 20\% | 80\% | 20\% | -12.00 EUR |
| 9 | 90\% | 10\% | 90\% | 10\% | -14.75 EUR |
| 10 | 100\% | 0\% | 100\% | 0\% | -17.50 EUR |

## A.1.2 Experimental Instructions

Preliminary remarks
You are participating in a study of choice behavior for the purpose of experimental economic research. During the experiment you and the other participants are asked to make decisions. In doing so, you can earn money. The resulting amount depends on your decisions and the decisions of the other participants.

The experiment will last approximately 150 minutes and consists of three parts. Prior to each of the three parts, you will receive detailed instructions. Note, that neither your decisions made in the first or second part nor the decisions made in the third part will have an influence on the respective other parts. Moreover, there are neither right nor wrong answers in any of the two parts.

Please note that the payment method distinguishes in all three parts. Details are mentioned in in the instructions under the headline "Payoff modalities".

Please read the following instructions carefully. In case you have any questions, please raise your hand.

## Part A: Risk (Part 1) \& Time Preferences (Part 2) ${ }^{1}$

## Part 1

You will be asked to take a series of decisions. In all decisions you receive two lists with 10 binary choices each. For each choice in every row you can choose between Option A and Option B. In total you have to take 20 decisions which are potential outcomes.

Whether one of these decisions is actually paid out is random (therefore "potential outcome"). At the end of the experiment, after you finished all three parts, the computer will draw a random decision and show the result on your screen. After that one experiment participant will be drawn by a lottery. This participant will be paid out. Further details are described in detail later.

Therefore one participant in the laboratory stands the chance to receive a payoff. The payoff depends upon chance and from your personal decision.

Decision situation

[^19]In the following the decision situation is described more detailed. One after the other you will see two lists on your screen. In each list you can choose in each row between two payout options.

The following example figure illustrates one decision situation.

| Option A |  | Choose an alternative in each row and press OK |  | Option B |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 15. | 12 E |  |  | 29 ¢ | 0,756 |
| 10 cs | mos | A | B | 1600 | nos\% |
| 2000 | soost | A | B | 2006 | s00\% |
| 30001 | rowt | A | B | 3000 | rown |
| 5005 | coss | A | B | 4005 | coss |
| S00\% | $500 \%$ | A | B | 50\% | 50\% 5 |
| 6005 | now | A | B | cous | 2000 |
| moos | soos | A | B | row | soos |
| 800\% | $200 \%$ | A | B | 500\% | 2004 |
| soos | 1006 | A | B | mos | 100\% |
| soens | -as | A | B | 1000\% | 00s |

Look at the first row of the decision situation with the alternatives A and B in the figure above: Option A has, with a probability of 10 Percent, a potential payoff of $15 €$ and with a probability of 90 Percent a payoff of $12 €$.

On the right side is Option B, which gives a potential payoff of $29 €$ (10 Percent probability) or a potential payoff of $0,75 €(90 \%$ probability $)$.

The difference between the payoffs of Option A is lower than for Option B. The potential payoffs in the experiment differ from this example.

The other choices are similar to the preceding one with the difference that the probabilities for the higher payoffs rise when you go down the list. In row 10 you just choose between the higher payoffs.

For each row you choose between the alternatives by clicking on the A or B. You have following possibilities:

- You can choose A for each row
- You can choose B for each row
- You can choose A for one or various rows and switch to B afterwards.

If you switch from A to B the program will also choose B for the following rows. You are able to change this manually. After you have picked an option in each of the 10 rows please confirm your choice with "OK".

The only difference in the second list is the varied payoff in Option A and Option B. Please choose again for one Option in each of the 10 rows.

## Payoff modalities

After you have played all the parts of the experiment, one participant will be determined by a lottery. This participant will receive his payoff from part 1 . The payoff modalities of part 2 and part 3 will be described in the following instructions.

The lottery is structured as follows: At the end of part 3 your screen will show you a button named "Lottery". If you click on the button the computer will randomly choose one of your 20 decisions. The potential payoff from this decision will be shown on your screen.

After this step, we will collect the table tennis balls which shows the number of your cabin. The participant with the randomly chosen cabin number will receive his payoff. If your number is chosen we will come and verify the payoff with you.

The real payoff for part 1 is made in cash at end of the experiment.

## Comprehension Questions

Prior to the decision rounds we kindly ask you to answer a few comprehension questions. These comprehension questions are intended to help you familiarize yourself with the decision situations and the payoff procedure. In case you have any questions, please raise your hand. part 1 of the experiment will begin once all participants have answered the comprehension questions correctly.

You will be asked to make a series of decisions similar to part 1. In this situation you will receive a list with 10 single choices. For this list you make decisions by choosing between Option A and Option B. Your choice represents a potential outcome.

Whether one of the decisions you make will actually be paid out to you depends on chance (therefore "potential outcome"). At the end of the experiment, after you finished all three parts, the computer will draw a random decision and show the result on your screen. The person who will actually receive the amount shown will then be drawn from among the experiment participants. This participant will be paid out. Note that for part 1 and part 2 different participants will be chosen. Further details are described in detail later.

In part 2 of the experiment, a participant in the laboratory again has the chance to receive a payout. This payment depends on the one hand on chance and on the other hand on your individual decisions.

## Decision situation

In the following the decision situation is described more detailed. You will see a list on your screen. In this list you can choose in each row between two payout options like in part 1. The two options offer a different amount that will be paid out and vary also in the payment period.

The following example figure illustrates one decision situation.

| Option A: <br> Amount in one week | Choose an alternative in each row and press OK$\square$ |  | Option B: <br> Amount in five weeks |
| :---: | :---: | :---: | :---: |
| $39 €$ | A | B | so $¢$ |
| 35 ¢ | A | B | $40 ¢$ |
| 31 ¢ | A | B | s0 $€$ |
| 27 ¢ | A | B | $40 ¢$ |
| 236 | A | B | $40 ¢$ |
| 19 € | A | B | $40 €$ |
| $15 ¢$ | A | B | 40 ¢ |
| $11 €$ | A | B | $40 €$ |
| 76 | A | B | $40 ¢$ |
| 36 | A | B | 406 |

The payout time described tells you when you will receive the potential payout..

You need to distinguish between the potential payout of option A to an earlier point in time and the potential payout of option B at a later stage. The amount of the earlier option decreases if you go down the list. The potential payoffs in the experiment differ from this example.

For each row you choose between the alternatives by clicking on the A or B. You have following possibilities:

- You can choose A for each row
- You can choose B for each row
- You can choose A for one or various rows and switch to B afterwards.

If you switch from A to B the program will also choose B for the following rows. You are able to change this manually. After you picked an option in each of the 10 rows please confirm your choice with "OK".

## Payoff modalities

At the end of the experiment, after you have completed all three parts, a participant will be chosen by lottery who will be paid for part 1 . A further draw will then determine a (different) participant who will be paid out for Part 2.

The lottery is structured as follows: At the end of part 3 your screen will show you a button named "Lottery". If you click on the button the computer will randomly choose one of your decisions in part 1 and one of your decisions in part 2 . The potential payoffs from these decisions will be shown on your screen.

After this step we will collect the table tennis balls which shows the number of your cabin. The first participant with the randomly chosen cabin number will receive his payoff for part 1. After this a second ball is randomly drawn. If your number is chosen we will come and verify the payoff with you.

If your number is chosen for part 2 you will receive your payoff after the selected period on your bank account. For this we need your account details, your name and your address after the experiment. The financial department of the University Duisburg-Essen will transfer the payment bank account ${ }^{2}$. For this purpose please stay in the laboratory after the experiment is finished.

[^20]The real payoff for part 1 is made in cash at end of the experiment.

## Comprehension Questions

Prior to the decision rounds of part 2 we kindly ask you to answer a few comprehension questions. These comprehension questions are intended to help you familiarize yourself with the decision situations and the payoff procedure. In case you have any questions, please raise your hand. Part 2 of the experiment will begin once all participants have answered the comprehension questions correctly.

## Part B: Health utilization under dynamic incentives

## Part 3 LowPrice (HighPrice)

In part 3 of the experiment all earnings are listed in Taler, the experimental currency units (ECU), where

## 1 Taler $=\mathbf{0 , 0 1 5}$ Euro

After finishing the experiment, your total earnings will be converted into Euro and paid out to you in cash. In part 3 all participants receive a payment.

Please read the following instructions carefully. In case you have any questions along the course of the experiment, please feel free to raise your hand. We will come to you.

Part 3 of the experiment lasts 52 rounds. At the start of each of these rounds you will receive an income of 50 Talers. In addition, a sickness can occur in each round, which can incur costs. There are three possible cases that can occur for each round::

- Healthy: No case of sickness occurs.
- Sickness A: The doctor's treatment costs 50 Taler. If you do not want to be treated by the doctor you bear the costs of 50 Taler as well.
- Sickness B: The doctor's treatment costs 50 Taler. If you do not want to be treated by the doctor you bear costs of 30 Taler.

The probability that no sickness will occur and that you will remain healthy is 60 percent in each round.

Both sicknesses A and B occur with a probability of 20 percent each. You can also imagine these probabilities like this: you roll a ten-sided die every round. If you have numbers 1 to 6 ,
you will remain healthy, if you have eyes 7 and 8 , you will experience sicknss A, and if you have eyes 9 and 10 , you will have sickness $B$.

Additionally, you are insured. The health insurance sets a deductible of 600 Taler in HighPricePeriods / 1150 Taler in HighPriceDed). This means that all costs in the 52 rounds ( 26 rounds in HighPricePeriod) up to an amount of 600 Taler ( $600 / 1150$ Taler) must be paid by yourself. All additional costs are paid by the insurance.

## Decision situation

At the beginning of each round you will be informed on your screen whether you are sick in this round or not.

If you do not get sick in a round, you will be credited with the periodic income of 50 Taler and you can continue the experiment by clicking "OK".

If you become sick in a round, you will also be credited with the periodic income of 50 Taler. You will also be informed on the screen whether it is sickness A or sickness B. You can then decide whether you want to have the sickness treated or not. You will also be shown the costs associated with your decision.

For each medical treatment you bear the costs yourself as long as the total sum of costs is below the deductible of 600 Taler ( 600 / 1150 Taler). All additional costs will be paid by the insurance.

If you do not want a medical treatment your costs will not be reported to the insurance and will not affect the deductible.

Example: In the first round, sickness A occurs. You will initially be credited with 50 Taler. If you choose the medical treatment you bear the 50 Taler yourself. The remaining treatment costs until the deductible is reached are reduced from 600 Taler (600/1150 Taler) to 550 Taler ( 550 / 1100 Taler). If you not want to be cured by a doctor you will bear the cost of 50 Taler. The deductible does not change and stays at 850 Taler ( $600 / 1150$ Taler).

## Information

After each round you will get three information:

1. Your total disposable income earned to date,
2. the total of your previous treatment costs at the doctor and
3. the remaining amount of treatment costs until the deductible is reached.

## Payoff modalities

After completing the experiment, your total income from part 3 will be converted into EUR and paid to you in cash. 1 Taler is 0.015 Euro.

First your secure profit in part 3 is shown on your screen. If you click on the button "Lottery" the computer will randomly choose one of your decisions in part 1 and one of your decisions in part 2. The potential payoffs from these decisions will be shown on your screen.

In the next step we will collect the table tennis balls that show you your cabin number and randomly draw two table tennis balls one after the other.

The participant with the cabin number that corresponds to the first table tennis ball drawn will receive the computer-determined decision from Part 1 . The participant with the cabin number that corresponds to the second table tennis ball drawn will receive the decision from part 2. If your number is chosen we will come and verify the payoff with you.

If your number is chosen for part 1 you will get this payment together with your profit from part 3.

If your number is chosen for part 2 you will receive your payoff after the selected period on your bank account. To do this, after the experiment we will note down your bank details, your name and your address and arrange for the transfer to be made by the finance department at the University of Duisburg-Essen. For this purpose, please stay in the laboratory for a short time after completing the experiment. ${ }^{3}$ You will receive the profits for part 3 immediately in cash.

## Comprehension Questions

Prior to the decision rounds of part 3 we kindly ask you to answer a few comprehension questions. These comprehension questions are intended to help you familiarize yourself with the decision situation and the payoff procedure. In case you have any questions, please raise your hand. Part 3 of the experiment will begin once all participants have answered the comprehension questions correctly.

## Part 3 (Neutral Frame)

In part 3 of the experiment all earnings are listed in Taler, the experimental currency units (ECU), where

## 1 Taler $=\mathbf{0 , 0 1 5}$ Euro

[^21]After finishing the experiment, your total earnings will be converted into Euro and paid out to you in cash. In part 3 all participants receive a payment.

Please read the following instructions carefully. In case you have any questions along the course of the experiment, please feel free to raise your hand. We will come to you.

Part 3 of the experiment lasts 52 rounds. At the start of each of these rounds you will receive an income of 50 Talers. In addition, a case of damage can occur in each round, which can incur costs. There are three possible cases that can occur for each round::

- No damage: No case of damage occurs. There are no costs for you.
- Damage A: Repairing the damage costs 50 Taler. If you do not have the damage repaired, you bear the costs of 50 Taler as well.
- Damage B: Repairing the damage costs 50 Taler. If you do not have the damage repaired, you bear costs of 30 Taler.

The probability that no case of damage will occur is 60 percent in each round.
Both cases of damage A and B occur with a probability of 20 percent each. You can also imagine these probabilities like this: you roll a ten-sided die every round. If you have numbers 1 to 6 , no case of damage occurs, if you have eyes 7 and 8 , you will experience damage A, and if you have eyes 9 and 10, you will have damage B.

Additionally, you are insured. The health insurance sets a deductible of 600 Taler, this means that all costs for damage repair in the 52 rounds up to an amount of 600 Taler must be paid by yourself. All additional costs are paid by the insurance, i.e. you do not bear any costs yourself for any further recoveries.

## Decision situation

At the beginning of each round you will be informed on your screen if there is a case of damage in this round or not.

If no case of damage occurs in a round, you will be credited with the periodic income of 50 Taler and you can continue the experiment by clicking "OK".

If a case of damage occurs in a round, you will also be credited with the periodic income of 50 Taler. You will also be informed on the screen whether it is damage A or damage B. You can then decide whether you want to have the damage repaired or not. You will also be shown the costs associated with your decision.

If you decide to repair the damage, you bear the costs yourself as long as the total sum of costs is below the deductible of 600 Taler. All additional costs will be paid by the insurance.

If you decide against repair your costs will not be reported to the insurance and will not affect the deductible.

Example: In the first round, damage A occurs. You will initially be credited with 50 Taler. If you have the damage repaired, you bear the 50 Taler yourself. The remaining costs to repair the damage until the deductible is reached are reduced from 600 Taler to 550 Taler. If you do not have the damage repaired, you will bear the cost of 50 Taler. The deductible does not change and stays at 600 Taler.

## Information

After each round you will get three information:
4. Your total disposable income earned to date,
5. the total of your previous costs to repair the damage and
6. the remaining amount of costs to repair the damage until the deductible is reached.

## Payoff modalities

After completing the experiment, your total income from part 3 will be converted into EUR and paid to you in cash. 1 Taler is 0.015 Euro.

First your secure profit in part 3 is shown on your screen. If you click on the button "Lottery" the computer will randomly choose one of your decisions in part 1 and one of your decisions in part 2. The potential payoffs from these decisions will be shown on your screen.

In the next step we will collect the table tennis balls that show you your cabin number and randomly draw two table tennis balls one after the other.

The participant with the cabin number that corresponds to the first table tennis ball drawn will receive the computer-determined decision from Part 1. The participant with the cabin number that corresponds to the second table tennis ball drawn will receive the decision from part 2. If your number is chosen we will come and verify the payoff with you.

If your number is chosen for part 1 you will get this payment together with your profit from part 3 .

If your number is chosen for part 2 you will receive your payoff after the selected period on your bank account. To do this, after the experiment we will note down your bank details, your name and your address and arrange for the transfer to be made by the finance department at the

University of Duisburg-Essen. For this purpose, please stay in the laboratory for a short time after completing the experiment. ${ }^{4}$ You will receive the profits for part 3 immediately in cash.

## Comprehension Questions

Prior to the decision rounds of part 3 we kindly ask you to answer a few comprehension questions. These comprehension questions are intended to help you familiarize yourself with the decision situation and the payoff procedure. In case you have any questions, please raise your hand. Part 3 of the experiment will begin once all participants have answered the comprehension questions correctly.

## A.1.3 Control Questions Ex-ante the Experiment

English translation of control questions for LowPrice (HighPrice):

1. What is the level of the deductible at the start of the experiment? Correct answer: 600 (1050)
2. What are the costs for additional treatment after you have incurred treatment costs equal to the level of the deductible? Answer options: $0 ; 20 ; 50$. Correct answer: 0
3. What is your remaining income in one round if you decide against treatment for illness B? Answer options: 0; 20; 50. Correct answer: 20
[^22]
## A. 2 Future Price and Expected Utilization

Figure A.2.1 is inspired by the intuitive model illustration in Aron-Dine et al. (2012) and adjusted to our parameters deductible level and number of periods, and definition of the future price. The upper panel in Figure A.2.1 illustrates that the expected future price, and thus the probability of reaching the deductible, depends on the number of periods left as well as the level of the deductible. The circle on the bottom left marks our LowPrice condition where a subject has the full 52 periods to reach the deductible of 600 . Here the probability of hitting the deductible is almost 100 percent, resulting in a future price of close to 0 . Future prices of hypothetical later entry dates, implying fewer periods to utilize the deductible, into the same deductible plan of 600 are located along the dashed line from this circle. Keeping the deductible constant, the future price of utilization increases when the time to utilize the same deductible decreases. Our HighPricePeriod condition, where subjects have only 26 periods to reach the deductible of 600 , also lies on this line and is marked with a diamond. Analogously, keeping the time to utilize constant at 52 periods, the future price increases with the level of the deductible. This is reflected in our HighPriceDed condition with a deductible of 1150, marked with a square. The connected line displays the future prices for different hypothetical entry dates into a 1150 deductible contract. The future price between the two HighPrice conditions is initially almost identical, which allows us to explore whether the channel that varies the future price affects utilization behavior differently. Our variation of the future price, while keeping the spot price constant, creates conditions akin to the "ideal experiment" proposed by Aron-Dine et al. (2015, p.726) in a simplified yet controlled setting. Hence, in the experiment we investigate the three specific contract combinations of deductible level and periods left to reach the deductible as depicted in Figure A.2.1. The lower part of Figure A.2.1 illustrates the expected number of treatments for sickness A and sickness B, for the three actual contracts as well as other hypothetical contracts.

Figure A.2.1: Future Prices and Expected Initial Utilization


## A. 3 Time Trend in HighPricePeriods

Table A.3.1: Probability to Treat Sickness B in HighPricePeriods

|  | Treatment | $p$-value |
| :--- | :---: | :---: |
| Period | -0.0711 |  |
|  |  | 0.004 |
| cons | -0.115 | 0.828 |
| lnsig2u | 1.275 | 0.138 |
| $N$ | 144 |  |
| rho | 0.782 |  |
| sigma_u | 1.892 |  |

Notes. This tables shows a random effect probit regression with participant's decision to seek treatment in the six cases of sickness B during 26 periods is dependent variable. ${ }^{*} p<0.1,{ }^{* *} p<0.05,{ }^{* * *} p<0.01$.

## A. 4 Heterogeneity in Utilization Behavior

Table A.4.1: Classification of Treatment Behavior by Period 26 per Condition

|  | Never treat B | Mixed Strategy | Always treat B | N |
| :--- | :---: | :---: | :---: | :---: |
| LowPrice | $9(18.75 \%)$ | $11(22.92 \%)$ | $28(58.33 \%)$ | 48 |
| LowPriceNoInfo | $8(16.67 \%)$ | $12(25.00 \%)$ | $28(58.33 \%)$ | 48 |
| LowPiceRev | $3(15.00 \%)$ | $5(25.00 \%)$ | $12(60.00 \%)$ | 20 |
| LowPriceNeutral | $4(8.51 \%)$ | $15(31.91 \%)$ | $28(59.57 \%)$ | 47 |
| HighPricePer | $10(41.67 \%)$ | $11(45.83 \%)$ | $3(12.50 \%)$ | 24 |
| HighPriceDed | $13(54.17 \%)$ | $10(41.67 \%)$ | $1(4.17 \%)$ | 24 |
| HighPriceDNoInfo | $11(45.83 \%)$ | $8(33.33 \%)$ | $5(20.83 \%)$ | 24 |

Notes. This table shows the classification of treatment behavior, indicated by the number of individuals, by period 26 for each condition. The respective percentages are provided in parentheses.

## A. 5 Classification of Behavioral Types

Figure A.5.1: Behavioral Types under Mixture of EUT and RDU


Figure A.5.2: Behavioral Types under EUT


Figure A.5.3: Behavioral Types under RDU


## A. 6 Average Number of Treatments after 52 Periods

Table A.6.1 Behavior by Period 52 (After 22 Sickness Cases; 9x A; 13xB)

|  | Avgerage number of <br> treatment choices | Treatment rate for <br> Sickness A | Treatment rate for <br> Sickness B |
| :--- | :---: | :---: | :---: |
| LowPrice | 18.38 | 0.98 | 0.73 |
| HighPriceDed | -5.11 | 0.94 | 0.15 |
| LowPriceNoInfo | 10.46 |  |  |
|  | -3.6 | 0.98 | 0.75 |
| HighPriceNoInfo | 18.54 | 0.93 | 0.23 |
| LowPriceReverse | -5.28 | 0.85 |  |
| LowPriceNeutral | 11.33 |  |  |

Notes. This table shows the average number of treatment cases by period 52. The treatment rate indicates share of respective sickness cases treated by all subjects.

## A. 7 Formal Description of Estimation of Individual Risk Preference Parameters

Assume that the utility of income from an experimental lottery choice task is defined by the following constant relative risk aversion (CRRA) specification:

$$
\begin{equation*}
U(x)=x^{(1-r)} /(1-r) \tag{1}
\end{equation*}
$$

where $x$ is the lottery prize and $r$ represents a coefficient that indicates the level of constant relative risk aversion. With this specification $r=0$ describes risk-neutrality, $r<0$ corresponds to risk-loving preferences, and $r>0$ corresponds to risk-averse preferences. Given an estimate of $r$ one can calculate the expected utility of a typical lottery $i$. If lottery $i$ has $j$ possible outcomes, the EU of the lottery is given by

$$
\begin{equation*}
E U_{i}=\sum_{j} p\left(x_{j}\right) U\left(x_{j}\right) \tag{2}
\end{equation*}
$$

Then, for each decision pair an index is calculated that indicates the difference in the expected utility of both lotteries in a decision pair. Formally,

$$
\begin{equation*}
\Delta E U=E U_{L}-E U_{R} \tag{3}
\end{equation*}
$$

where $E U_{L}$ is the "left" lottery and $E U_{R}$ is the "right" lottery in a decision pair as presented to subjects. The function that links the latent index in (3) to observed choice behavior is the cumulative density function (cdf) of the univariate normal distribution $\Phi(\cdot)$, resulting in a probit model. The probability of choosing the "right" lottery can be written

$$
\begin{equation*}
\operatorname{prob}(\text { choose } R)=\Phi(\Delta E U) . \tag{4}
\end{equation*}
$$

Thus the latent index in (3) is linked to the observed choices by making the assumption that lottery R is chosen, when the $\Delta E U>0.5$.

This basic approach can be extended in several ways. An important addition is accounting for behavioral errors. The structural probit model cannot predict individual decision making with certainty. Decision makers may deviate from their true underlying preferences for a variety of reasons. Behavioral error specifications can account for various error sources, ranging from random deviations due to attention lapses to systematic violations related to the psychology of perception and judgment. A particularly influential behavioral error specification is due to Fechner (1860). Its application to the evaluation of risky prospects was popularized by Hey and Orme (1994). The inclusion of the Fechner error specification expands the latent index in (3) to

$$
\begin{equation*}
\Delta E U=\left(E U_{L}-E U_{R}\right) / \mu, \tag{5}
\end{equation*}
$$

where the new parameter $\mu$ allows the otherwise deterministic EUT model to account for deviations from the underlying preference structure.

Wilcox $(2008,2011)$ suggests an additional characterization of behavioral errors, called "contextual utilty." The intuition behind contextual utility originates from psychological
experiments on signal detection and stimulus discrimination. These studies discovered that errors became more likely as the range of possible stimuli increase. Contextual utility respects this observation, by assuming that evaluative errors increase with the perceived range of outcomes. Econometrically, this implies that the standard deviation of the behavioral error is proportional to the range of utilities of the outcomes in a lottery pair. The contextual error specification is given by

$$
\begin{equation*}
\Delta E U=\left(E U_{L}-E U_{R} / v\right) / \mu, \tag{6}
\end{equation*}
$$

where the new parameter $v$ is defined as the maximum utility over all outcomes minus the minimum utility over all outcomes in the lottery pair, i.e., over the context of that pair. This specification has a normalizing effect on the latent index, which then remains in the unit interval. ${ }^{1}$

Once the parameters of interest are defined, structural estimation can be undertaken using the procedures explained by Andersen, Harrison, Lau and Rutström (2008) and Harrison and Rutström (2008).

People may not necessarily behave as if given probabilities affect their lottery evaluations with objective values. Instead, they may distort these probabilities using a process that can be described by attaching subjective weights to probabilities. The Rank Dependent Utility (RDU) model, due to Quiggin (1982) derives probability weights from the entire distribution over ranked outcomes, not from individual probabilities, and avoids any theoretical violations of first-order stochastic dominance. The resulting decision weights reflect subjective distortions of objective probabilities.

The RDU model nests the EUT model, and requires the introduction of a probability weighting function. A variety of weighting functions have been proposed in the literature, primarily by Quiggin (1982), Tversky and Kahneman (1992) and Prelec (1998). Prelec (1998) contributes a flexible two parameter specification of probability weighting:

$$
\begin{equation*}
\omega(p)=\exp \left[-\eta(-\ln p)^{\phi}\right], \tag{7}
\end{equation*}
$$

[^23]with $\eta>0$ and $\phi>0$. This weighting function is derived from several axioms that reflect apparent regularities of probability weighting, and requires the estimation of two additional parameters $\eta$ and $\phi$.

Another special case of RDU, due to Yaari (1987) and known as Dual Theory, assumes that all of the risk premium is due to probability weighting, and that there is a linear utility function such that $r=0$ and hence that $U^{\prime \prime}=0$. In this case a risk premium is generated entirely by "pessimistic probability weighting" with respect to better-ranked outcomes. There is very little empirical evidence to support the use of Dual Theory (see Harrison and Swarthout, 2023).

## References

Fechner, G.T. (1860). Elemente der Psychophysik. Leipzig, Germany: Breitkopf and Härtel.
Hey, J. D. \& Orme, C. (1994). Investigating Generalizations of Expected Utility Theory using Experimental Data. Econometrica, 62(6), 1291-1326.

Wilcox, N. T. (2008). Stochastic Models for Binary Discrete Choice under Risk: A Critical Primer and Econometric Comparison. In: J. Cox and G. W. Harrison (Eds), Risk Aversion in Experiments. Bingley, UK: Emerald, Research in Experimental Economics, volume 12.

Wilcox, N. T. (2011). 'Stochastically More Risk Averse': A Contextual Theory of Stochastic Discrete Choice Under Risk. Journal of Econometrics, 162, 89-104.

Yaari, M. E. (1987). The dual theory of choice under risk. Econometrica, 55(1), 95-115

## A. 8 Welfare Effects for RDU and EUT

Figure A.8.1 displays results for ECS in the left panel, and results for Efficiency in the right panel. We present results assuming a mixture model of EUT and RDU risk preferences in red, and assuming EUT risk preferences in blue We present results assuming a RDU risk preferences in red, and assuming EUT risk preferences in blue. Under RDU we see a large number of small welfare losses, yet all losses are below the level of $€ 2$. On the other hand, under EUT we find many tiny ECS welfare losses, but a long tail of larger losses. These findings for ECS translate into a single mode for Efficiency under RDU at high levels close to $100 \%$, and a slight tail of lower efficiency levels. For EUT we observe two modes: a significant number of subjects around 50 to $70 \%$ Efficiency, and then some subjects with between 0 and 25\% Efficiency. These results point to more individuals making mistakes that were welfare costly under EUT, to the point where their Efficiency drops well below $50 \%$ in many cases. Under EUT we find that more than two thirds of the subjects are classified forward-looking. Hence, these costly mistakes appear to originate from a few subjects making very large mistakes.

Figure A.8.1: Total Foregone Welfare


We can explore these welfare results by examining the "marginal effects" of treatments
or demographics on welfare distributions. In this case we consider the task and subject characteristics shown in Figure A.8.2 for RDU and Figure A.8.3 for EUT. The diamond symbol is the point estimate of the marginal effect, and the bars either side of that symbol show the $95 \%$ confidence interval.

For both EUT and RDU we find a significant effect of the low probability of reaching a deductible leading to a welfare effect. Under RDU this effect is positive, but under EUT it is negative. The loss is particularly strong under EUT, and appears to be the driving factor behind the large tail of losses in ECS noted earlier under EUT for Figure 4. We also identify welfare effects that are specific with respect to the variations in our robustness conditions. Under RDU we find that there is a tendency for extra information on the remaining level of the deductible to lower welfare. Under EUT we find only a tiny negative effect of extra information. For the abstract insurance context, we find a higher welfare under RDU and only a tiny negative effect on welfare under EUT.

Figures 5 and 6 also show that there are welfare effects which depend on individual subject characteristics. Under RDU we find a significant welfare reduction for women. In contrast, under EUT there is a significant increase in welfare for women. Studying in the field of economics and having a private health insurance increase welfare under both RDU and EUT. The effect of statutory health insurance on welfare differs depending on the underlying model of risk preferences. Under RDU having statutory health insurance has a significant negative effect on welfare while it has a significant positive effect under EUT.

Figure A.8.2: Marginal Effects on Efficiency under RDU
Marginal average effects from a Fractional regression model


Task Characteristics
Low probability of reaching deductible
Extra information
Abstract context

## Subject Characteristics

Female

- Age in years (over 17)

Field of study is Economics

- Has private health insurance

Has own statutory health plan
$-6 \%-5 \%-4 \%-3 \%-2 \%-1 \% ~ 0 \% ~ 1 \% ~ 2 \% ~ 3 \% ~ 4 \% ~$
Percentage Point Effect on Efficiency

Figure A.8.3: Marginal Effects on Efficiency under EUT
Marginal average effects from a Fractional regression model



[^0]:    * Department of Risk Management \& Insurance and Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, USA. Harrison is also affiliated with the School of Economics, University of Cape Town. E-mail contact: gharrison@gsu.edu.
    ${ }^{\dagger}$ Heinrich Heine University Düsseldorf, Institute for Health Services Research and Health Economics, Moorenstr.5, 40225, Düsseldorf, Germany \& CINCH - National Health Economics Research Center, Essen; email: johann.han@uni-due.de, nadja.kairies-schwarz@hhu.de (corresponding author).
    \$We thank Daniel Avdic, Friedrich Breyer, Amitabh Chandra, Simon Decker, Ben Handel, Mathias Kifmann, Johanna Kokot, Jimmy Martinez Correa, Justin Sydnor, Christian Waibel, as well as participants at the dggö annual meeting of the German Health Economics Association 2017 in Basel, Young Economists Meeting in Brno, HESG Meeting in Aberdeen, 4th EuHEA PhD conference in Lausanne, the German Health Economics Association workshop on "Allocation and Distribution" in Hamburg, the 5th Workshop in Behavioral and Experimental Health Economics in Atlanta, the German Economic Association 2018 in Freiburg, the SocioScientific Committee of the German Economic Association 2019 in Essen, the CEAR/MRIC Behavioral Insurance Workshop 2022 in Munich. Financial support provided by the Bundesministerium für Bildung und Forschung (Federal Ministry of Education and Research) is gratefully acknowledged.

[^1]:    ${ }^{1}$ More complex insurance contracts might also use coinsurance where the insured pays a fraction of the costs above any deductible. We focus exclusively on deductibles, assuming full insurance beyond the payment of the deductible. Pauly (1968), Zeckhauser (1970), and Arrow (1971) are classic references on the mitigation of moral hazard in insurance contract design.
    ${ }^{2}$ See, for example, Keeler and Rolph (1988); Aron-Dine et al. (2013) for the RAND Health Insurance Experiment, and Finkelstein et al. (2012) for the Oregon Health Insurance Experiment.
    ${ }^{3}$ See, for example, Cardon and Hendel (2001), Kowalski (2015), Aron-Dine et al. (2015), Brot-Goldberg et al. (2017), Guo and Zhang (2019), Klein et al. (2022) or Johansson et al. (2023) for health insurance plans, and Einav et al. (2015), Dalton et al. (2020) or Abaluck et al. (2018) for Medicare Part D plans. Earlier theoretical contributions addressing the issue include Keeler et al. (1977) and Ellis (1986).

[^2]:    ${ }^{4}$ See, for example, Schram and Sonnemans (2011), Krieger and Felder (2013), Kairies-Schwarz et al. (2017)
    Mimra et al. (2020), Samek and Sydnor (2020), Biener and Zou (2022) or Hermanns et al. (2023).
    ${ }^{5}$ Some evidence suggests that many individuals do not completely understand them. The most immediate evidence is from the choice of dominated strategies, in particular when they are transparent, as in Bhargava et al. (2017), Biener and Zou (2022) or Samek and Sydnor (2020). For a broader discussion of insurance literacy see Harrison et al. (2022).
    ${ }^{6}$ Several studies suggest that providing individuals with information, or simplifying the decision process, can indeed affect decision outcomes. See Kling et al. (2012) for Medicare Part D plan choices, Hastings et al. (2008)

[^3]:    ${ }^{7}$ We used this order because we did not want income effects from the utilization part to affect the elicitation of preferences. Since the payments for the elicitation part were randomly determined after the whole experiment was concluded, this concern did not arise with the given order. We controlled for potential order effects by reversing the order of both parts in one condition.
    ${ }^{8}$ The general design was inspired by a dynamic model outlined in Aron-Dine et al. (2012) and is similar to Einav et al. (2015) in the Medicare D prescription drug context.
    ${ }^{9}$ The conversion rate at the time of the experiment was $1 \mathrm{ECU}=0.015$ EUR.
    ${ }^{10}$ An obvious extension is to consider the compound lottery in which one sick or not, but 'treatment' is needed to identify, through medical tests and diagnosis, which sickness it is.

[^4]:    ${ }^{11}$ Since we construct the future price under the assumption that all sicknesses are treated, it only matters if subjects are sick or healthy, not the severity of illness. Hence we can retrieve the future prices from the cumulative distribution function of the binomial distribution where the number of trials equals the number of periods and the number of successes is the number of treatments needed to hit the deductible, e.g. 12 for a deductible of 600. The probability of being sick, regardless of severity, is $p=0.4$.

[^5]:    ${ }^{12}$ To calculate an expected number of visits, we assume a basic understanding of the dynamic incentives in the sense that it is reasonable to reduce the deductible. Hence, this individual would choose treat sickness A and visit the doctor in this case. If this individual had a preference for not treating A or would use a mixed strategy, the expected number of treatments would be even lower.
    ${ }^{13}$ Here, we multiply the probability of occurrence of sickness A with the number of number of the initial 26 periods, i.e. Expected\#treatments myopic $=\operatorname{Pr}($ SicknessA $) \cdot 26=0.2 \cdot 26=5.2$

[^6]:    ${ }^{14}$ For LowPrice in the beginning the expected number of treatments for a completely forward-looking individual is: Expected\#treatments LowPrice $_{\text {Low }}=\operatorname{Pr}($ SicknessA $) \cdot 26+\operatorname{Pr}($ HitAB $) \cdot \operatorname{Pr}($ SicknessB $) \cdot 26=0.2 \cdot 26+$ $0.997 \cdot 0.2 \cdot 26=10.4$. Thus, we expect a forward looking individual to base his treatment decision for B on the initial probability to hit the deductible if it would treat both cases. $\operatorname{Pr}(H i t A B)=0.326$ in HighPriceDed and $\operatorname{Pr}($ Hit AB $)=0.3$ in HighPricePeriod.
    ${ }^{15}$ The relationship between the expected future price and expected utilization behavior is presented in more detail in Appendix A.2.
    ${ }^{16}$ These are the dominant strategies when assuming risk neutrality and a preference for treating sickness A over not treating it, which signals a basic understanding of the dynamic structure.
    ${ }^{17}$ The forward-looking type recognizes that the probability to hit the deductible is smaller than 50 percent.

[^7]:    ${ }^{18}$ Profits in Experimental Currency Unit (ECU) with $1 \mathrm{ECU}=0.015$ EUR.

[^8]:    ${ }^{19}$ The first part consisted of the 20 choices over risky lotteries referred to earlier, as well as 10 choices over time-dated amounts of money. These 10 choices are not relevant to the decisions within a 90 -minute session, and are not discussed further. However, two subjects were randomly selected and were each paid in the first part for one of their 30 decisions. One individual was paid for one ex post randomly determined choice over risky lotteries and one individual for one ex post randomly determined choice over time-dated amounts of money. On average we had about 21 subjects within each session.

[^9]:    ${ }^{20}$ We run a random effects probit regression with the decision to seek treatment or not in HighPricePeriod as the dependent variable and the respective Periods of sickness B as independent variable. The coefficient is negative and significant on a level of 0.05. See Table A.3.1 in Appendix A.3.

[^10]:    ${ }^{21}$ In a natural setting time preferences would play a role as well. In our experimental design, given that the time horizon was about 1 hour, it is plausible to assume that there is no time preference within the session.

[^11]:    ${ }^{22}$ We aggregate the LowPrice treatments here. As shown above, we do not find significant differences in aggregate behavior treatment behavior between these conditions. For details, see Appendix A. 4 Table A.4.1.

[^12]:    ${ }^{23}$ Thus, at least 70 percent of the individuals stick to one strategy: they always or never treat sickness B in one of the treatments with a low price. Taking a closer look at the average treatment rates of the subjects who do not stick to one of the two strategies, we do not find a clear pattern in their behavior over time. Individual decision patterns of these subjects show that few individuals decide against treating sickness $B$ at its first occurrence, and always treat later, which could be attributed to learning. However, for the majority this is not a plausible behavioral explanation.
    ${ }^{24}$ For the horizon we are considering within the experiment we assume that individuals have no discounting at all. Accounting for risk preferences should therefore be sufficient to derive optimal strategies. Yet, a natural expansion of the experimental design would be to extend the horizon of the experiment and hence also account for individual time preferences.

[^13]:    ${ }^{25}$ Gao et al. (2023; §3.1) evaluate this issue by comparing inferences about welfare when each subject made 80 binary choices with inferences about welfare when just 20 choices per subject were selected at random. They find an acceptably high correlation of inferences about individual welfare, precisely to guide experimental designs in answering the question of "how many binary choces are needed" to generate reliable welfare evaluations. Obvioulsy, more (informative) choices are always better than fewer, but in practice it is valuable to have guidance on the number of choices that are likely to be sufficient for reliable welfare evaluation.
    ${ }^{26}$ We could also estimate a Dual Theory (DT) model in which the utility function is assumed to be linear, and the risk premium for an individual is generated entirely by their estimated probability weighting function. DT is also nested in RDU. However, we have never found systematic evidence that any noticeable fraction of individual subjects exhibit DT behavior. And DT is not regarded as normatively attractive by anyone, as far as we are aware (see Harrison and Swarthout, 2023).

[^14]:    ${ }^{27}$ For a formal description of the estimation of individual risk preference parameters see Appendix A.7.

[^15]:    ${ }^{28}$ Treat only the mild sickness B and treat nothing are other pure strategies but they are statistically dominated. For simplicity and as mentioned in Section 2.3., we do not look at mixed strategies.
    ${ }^{29}$ From the perspective of period 0 and with $P(O O P)_{i}^{j}$ being the probability to treat out of pocket in period $j$ given strategy $i$, the expected costs in period 1 would be $\mathrm{P}(O O P)_{A B}^{1} \cdot P(A \cup B) \cdot C($ Treat $)$ for strategy "treat both A and $\mathrm{B} "($ abbreviated as $A B)$ and $P(O O P)_{A}^{1} \cdot P(A) \cdot C($ Treat $)+\mathrm{P}(\mathrm{B}) \cdot \mathrm{C}($ NotTreat $)$ for the strategy to treat "A only" (A), where $C$ are the costs for treating (Treat) or not treating (NotTreat) the respective sickness. To begin we sum up the expected cost for every upcoming period and yield the cumulated utility for these two strategies from the perspective of period 0 . Then we continue with the perspective of period 1 , and do the same calculations. To get the expected utilities we just replace the cost function with utility functions.

[^16]:    ${ }^{30}$ Given our experimental design, treatment for sickness B is socially inefficient: the treatment cost in case of sickness B in a given period is 50 while the opportunity cost is only 30 . When an individual understands that treating sickness B might be individually welfare optimizing in the low price treatments and treats sickness B , this results in a periodic social welfare loss of 20. While we cannot rule out the hypothesis that individuals who did not treat sickness B were in fact forward-looking, but were acting as social welfare maximizers, the design was such that there was no individual or group that suffered potential welfare losses from treating sickness B. We therefore believe that this motive did not play a pivotal role in the treatment decisions of individuals.
    ${ }^{31}$ For a direct welfare comparison of EUT and RDU see Appendix A.8.

[^17]:    ${ }^{32}$ The exception is if we assume that agents have some unbiased distribution of expected health care costs and then apply the Reduction of Compound Lotteries assumption to reduce that compound risk to a scalar. This is assumed under EUT and RDU, as it happens, ad effectively "assumed away" the possible effects of uncertainty. It is also widely assumed in most empirical evaluations of health insurance using field data: see Harrison (2024) for a detailed review.

[^18]:    ${ }^{33}$ The ex ante control questions for the LowPrice (HighPrice) condition were as follows: What is the level of the deductible at the start of the experiment? Correct answer: 600 (1050). What are the costs for additional treatment after you have incurred treatment costs equal to the level of the deductible? Answer options: 0;20;50. Correct answer: 0 . What is your remaining income in one round if you decide against treatment for illness B? Answer options: 0; 20; 50. Correct answer: 20.

[^19]:    ${ }^{1}$ Although we do not discuss the results of this part, we have included them for completeness.

[^20]:    ${ }^{2}$ We need your address to arrange the payment by the University Duisburg-Essen. We will not use your personal information for any other purpose.

[^21]:    ${ }^{3}$ We need your address to arrange the payment by the University Duisburg-Essen. We will not use your personal information for any other purpose.

[^22]:    ${ }^{4}$ We need your address to arrange the payment by the University Duisburg-Essen. We will not use your personal information for any other purpose.

[^23]:    ${ }^{1}$ The contextual error specification is particularly parsimonious, since the parameter $v$ is defined by data, so that no additional parameter estimation is required. The specification also allows for inferences regarding "stochastically more risk averse" relationships. The latter refers to a stochastic notion of the familiar Arrow-Pratt metric of risk aversion. A stochastically risk averse subject is "on average" risk averse, but the metric is flexible enough to deal with choices that deviate from the subject's general risk aversion. With the latent index remaining within the bounds of the unit interval, one can compare the stochastic risk aversion of subjects who choose in dramatically different decision contexts (i.e., who face lotteries with very different prizes).

