WORKING PAPERS

N° 1449

June 2023

# “Measuring regret theory in the health and financial domain”

Henrik Andersson, Henrik Scholtz and Jiakun Zheng

Toulouse School of Economics

# Measuring regret theory in the health and financial domain

Henrik Andersson[*] Henrik Scholtz[†] Jiakun Zheng[‡]

June 22, 2023

**Abstract**

This paper applies an experimental design developed by Bleichrodt et al. (2010) to test the key assumption of original regret theory (Loomes and Sugden, 1982): convexity of the regret function. We elicit preferences for financial and health outcomes for about 1,000 subjects, yielding some evidence of minor differences between financial domain and health domain. While aggregate results seem to support regret theory at first sight, individual-level analyses show that the majority of subjects violate the predictions of regret theory with a convex regret function. Our results thus challenge the predictive accuracy of regret theory as a descriptive theory of decision-making under risk.

**Keywords:** Original regret theory; Decision under uncertainty; Utility measurement; Online experiments.

**JEL Classification Numbers:** C91, D81

[*]Toulouse School of Economics, University of Toulouse Capitole, Toulouse, France. Email (🖂): henrik.andersson@tse-fr.eu.

[†]University of Zurich, Department of Economics, Zürich, Switzerland. Email (🖂): henrik.scholtz@econ.uzh.ch.

[‡]School of Finance at Renmin University of China, China Financial Policy Research Center and China Insurance Institute, Beijing, China. Email (🖂): jiakun.zheng@ruc.edu.cn.

# 1 Introduction

For more than half a century, social scientists have proposed and developed theories of decision making under risk that challenge the predominant model of expected utility theory (EUT). Much of this research was driven by experimental violations of EUT, notably including the common consequence Allais paradox (Allais, 1953), the common ratio effect, and the reflection effect (Kahneman and Tversky, 1979). Among the proposed alternative theories,[1] regret theory (RT) developed by Bell (1982) and Loomes and Sugden (1982) is still an active field of both theoretical and experimental research.[2] The psychological underpinning of RT is that decision makers feel regret, a negative emotion, when realizing that they could have been better off had they behaved differently. They are thus motivated to minimize the possible future regret when making decisions (Zeelenberg, 1999). RT can accommodate some commonly observed violations of EUT under the assumption that people are disproportionately averse to large regrets. This notion of regret aversion is linked to a convexity property of the regret function, as initially pointed out by Loomes and Sugden (1982, 1987a). In contrast, if the regret function is linear, RT is isomorphic to EUT and thus yields identical predictions. If the regret function is concave, predictions are generally at odds with observed experimental and empirical behavior (Loomes and Sugden, 1982). Therefore, one way to directly test RT is to test the convexity of the regret function.

Despite the ongoing theoretical and empirical interest in RT, almost no effort has been made to quantitatively measure it.[3] To the best of our knowledge, the only two studies specifically designed to test the convexity of the regret function are Bleichrodt, Cillo, and Diecidue (2010) (hereafter referred to as "BCD") and Somasundaram and Diecidue (2017).[4] The parameters of Kahneman and Tversky (1979)'s prospect theory, on the other hand, have been measured in dozens of studies (e.g., Abdellaoui et al., 2007), making comparisons across studies possible. The literature on RT would likewise benefit from more substantial evidence regarding the convexity of the regret function and the predictive

[1] Starmer (2000) provides a review of the development of non-expected utility theories.

[2] See also Bleichrodt and Wakker (2015) for a more comprehensive review of RT and its applications.

[3] A different way of testing RT is to design choice situations in which RT predicts a particular type of behavior that is opposed to the predictions of most other common theories of decision-making under risk. This non-parametric approach was introduced by Loomes and Sugden (1987b) and Loomes (1988). However, it is incapable of detecting regret if its effect is only of second order. Loewenfeld and Zheng (2023) recently overcome these issues and develop a test that lets subjects choose between lotteries with identical marginal distributions but different correlational structures of payoffs. In their setup, RT and salience theory (Bordalo et al., 2012) predict a clear preference for one lottery while according to other theories, notably EUT and prospect theory, subjects should be indifferent. The experimental evidence they present is clearly not in favor of RT. If anything, it would be consistent with subjects having a concave regret function.

[4] Somasundaram and Diecidue (2017) replicated BCD's findings with a student sample and studied the impact of immediate choice feedback on risk and regret attitudes. We will discuss the main differences between our experiment and BCD's (and thus also Somasundaram and Diecidue's) in section 7.

accuracy of RT in cases where its predictions strongly differ from those of other theories of decision-making under risk.

The goal of this study is twofold. First, we aim to substantiate or refute the results of BCD in favor of RT using a much larger and arguably more representative sample. Secondly, we test whether risk and regret attitudes differ systematically between financial and health outcomes. Since health decisions are particularly important, complex and generally hard to undo, decision makers are more likely to anticipate regret and act upon it in this domain (Zeelenberg, 1999). Brewer et al. (2016) and Becerra Pérez et al. (2016) summarize a large body of evidence on the importance of regret for health decisions, yet RT has only rarely been applied in this research. One exception is Smith (1996) who applies RT to treatment decisions for chronic diseases. More recently, Zheng (2021) investigates willingness to pay for health risk reductions in a framework of anticipated regret, and shows that disproportionate aversion to large regrets exaggerates willingness to pay estimates. However, it remains unclear whether individuals hold similar regret attitudes towards health and financial risks.

Closely following BCD's elicitation method, we conducted an online survey study with a total of 1,004 MTurk participants. After entering the survey, subjects were informed of which type of risks they were going to face during the survey. This could be either financial or health risks, depending on the treatment group they were assigned to. Specifically, subjects were assigned to a health scenario (H), an unincentivized financial scenario (U), or an incentivized financial scenario (I) with equal probabilities. On an aggregate level, just as BCD, we find some evidence supporting RT-conforming preferences. Taking a closer look at individual-level preferences, however, the majority of our subjects do not reveal preferences in line with RT under the assumption of a convex regret function. Using different measures of curvature of the regret function, we show that the behavior of a substantial share of subjects is more compatible with a concave than a convex regret function and violates predictions of RT in a specifically designed lottery choice.[5] We extend the analysis from incentivized financial outcomes to unincentivized financial outcomes and (hypothetical) health outcomes and find only minor differences in preferences between these three scenarios. Seen in the context of other recent tests of regret theory (e.g., Loewenfeld and Zheng, 2023), our estimation results challenge the validity of RT as a descriptive theory of choice under risk.

The remainder of this paper is organized as follows. Section 2 introduces original regret theory that is to be quantitatively measured; section 3 describes the measurement method initially adopted by BCD; sections 4 and 5 present the design and implementation of our online experiment and information on our subjects; section 6 gives our empirical results;

[5] According to Herweg and Müller (2021), original regret theory is just a special case of Bordalo et al. (2012)'s salience theory. In particular, there exists a specification of the value function and the salience function such that standard smooth salience theory is behaviorally equivalent to original regret theory with a power regret function (see proposition 1 of Herweg and Müller, 2021). So one should be cautious in relating our test results to Bordalo et al.'s salience theory. But it is worth noting that recent studies by Alós-Ferrer and Ritschel (2022) and Loewenfeld and Zheng (2023) both reject salience theory.

section 7 discusses our findings in comparison with BCD and the last section offers a conclusion.

## 2 Regret theory

In regret theory (RT), uncertainty is characterized by states of nature (Savage, 1951). Let $\mathcal{S}$ denote the set of states of nature. Subsets of $\mathcal{S}$ are called events. A probability measure $\mathcal{P}$ is given over the set of events. An act is a function from $\mathcal{S}$ to the set of outcomes. Like Bleichrodt, Cillo, and Diecidue (2010) (BCD), we restrict our attention to binary acts. We denote an act that leads to event $E$ yielding outcome $x$ with probability $p$ and to its complementary event $E^C$ yielding outcome $y$ with probability $1-p$ by $x_py$. In principle, it is not necessary to restrict consequences to be real numbers. They could also be non-monetary outcomes such as health status. However, in the following analysis, as in most applications of RT, all outcomes are represented by real numbers over which preferences are defined.

When making a choice, the decision maker in RT compares the outcome that her choice produces in a given state of nature act to the outcome that would have obtained had she made a different choice. A preference relation $\succeq$ is given over the set of binary acts. The conventional notations $\succ$ and $\sim$ are used to denote strict preference and indifference respectively. We assume that higher outcomes are preferred to lower outcomes. In particular, we assume that people prefer more money over less, and prefer to live longer rather than shorter.

RT in its general form, as defined by Loomes and Sugden (1987a), assumes that a preference relation over the set of binary acts satisfies

$$x_py \succeq v_pw \iff p\Psi(x,v) + (1-p)\Psi(y,w) \geq 0. \tag{2.1}$$

$\Psi(x,y)$, the regret function, is a real-valued function which is unique up to multiplication with a positive scalar, strictly increasing in its first argument, skew-symmetric, and convex in the sense that $\forall x > y > z, \Psi(x,z) > \Psi(x,y) + \Psi(y,z)$. Convexity of the regret function is necessary for RT to accommodate common violations of EUT. This property is also referred to as regret aversion. The ultimate goal of the elicitation procedure introduced by BCD is therefore to map out the regret function in order to test whether individuals indeed exhibit behavior that is consistent with convex regret functions.

In the following, we work with a less general version of the regret function, i.e.,

$$\Psi(x,y) = Q(U(x) - U(y)). \tag{2.2}$$

This regret theory specification, often referred as the original regret theory (Loomes and Sugden, 1982), disentangles standard choiceless utility, embodied by function $U(\cdot)$, and "behavioral" regret utility, embodied by function $Q(\cdot)$. The function $U(\cdot)$ is unique up to positive affine transformations. The regret function $Q$ evaluates the distance (in terms of utility) between the realized outcome $x$ and the outcome $y$ that would have been realized in the given state if the alternative act had been chosen. $Q$ inherits its major properties from $\Psi$. That is, $Q$ is an odd function and supposedly convex. Just as $\Psi$, $Q$ is only unique up to scale.

# 3 Measurement method

## 3.1 Utility

We start by eliciting a $k$-standard sequence of outcomes - that is, a sequence of length $k$ of outcomes with equal utility distance between them - using the tradeoff method (Wakker and Deneffe, 1996).[6] Formally, a standard sequence satisfies $U(X_{j+1}) - U(X_j) = U(X_j) - U(X_{j-1}),\ \forall j \in \{1, ..., k-1\}$. We select a starting outcome $X_0$, a probability $p \in (0,1)$ and two gauge outcomes $G$ and $g$ with $G > g$.[7] The goal is then to determine the outcome $X_1$ for which an individual is indifferent between $X_{1_p}g$ and $X_{0_p}G$. Using the regret function introduced above and its symmetry, this indifference condition can be expressed as

$$pQ(U(X_1) - U(X_0)) + (1-p)Q(U(g) - U(G)) = 0 \tag{3.1}$$

or equivalently

$$Q(U(X_1) - U(X_0)) = \frac{1-p}{p} Q(U(G) - U(g)). \tag{3.2}$$

In a similar manner, we find the outcome $X_2$ for which each individual is indifferent between $X_{2_p}g$ and $X_{1_p}G$. The corresponding indifference condition after rearranging is

$$Q(U(X_2) - U(X_1)) = \frac{1-p}{p} Q(U(G) - U(g)). \tag{3.3}$$

We therefore have $Q(U(X_2) - U(X_1)) = Q(U(X_1) - U(X_0))$. Monotonicity of $Q$ implies that $U(X_2) - U(X_1) = U(X_1) - U(X_0)$. By repeating the same process $k$ times and finding the respective values of $X_{j+1}$ that lead to $X_{j+1_p}g \sim X_{j_p}G$, we can elicit a $k$-standard sequence $\{X_1, ..., X_k\}$ where outcomes are equally spaced in terms of utility. Since $U$ is only unique up to a positive affine transformation, the value of two outcomes can be chosen arbitrarily so long as their relative ranking is maintained. We set $U(X_0) = 0$ and $U(X_k) = 1$, thereby standardizing the elicited part of the utility function to range $[0, 1]$. This standardization implies $U(X_j) = j/k$. Following BCD, we choose $k = 5$, thus eliciting five (excluding $X_0$) outcomes with an equal utility distance of 0.2 between them. We confirmed in our pilot sessions that this value yields a sufficient number of observations

[6] Recent work by Richard and Baudin (2020) shows that the trade-off method can induce systematic bias in the estimation of the utility function. However, how the systematic bias pointed out in this paper would affect the estimation of regret theory is unclear and this question should be further investigated.

[7] Including the starting outcome $X_0$, the sequence then has length $k+1$.

per subject to classify the shape of the utility function without lengthening the survey to an unreasonable extent.

### 3.2 Regret attitude

Using the standard sequence elicited before, we choose a probability $p$ and two elements from the standard sequence, say $X_j$ and $X_{j+1}$, to determine the value $Z$ for which a given subject is indifferent between $X_{j+1_p}X_0$ and $X_{j_p}Z$. While BCD arbitrarily chose $X_j = X_3$ and $X_{j+1} = X_4$, we see no reason to prefer one particular pair of outcomes from the standard sequence and therefore randomized $X_j$ from the set $\{X_2, X_3, X_4\}$.[8]

The following indifference condition can be derived from the aforementioned pair of acts:

$$pQ(U(X_{j+1}) - U(X_j)) + (1-p)Q(U(X_0) - U(Z)) = 0. \tag{3.4}$$

Since $U(X_0) = 0$ and Q is symmetric, we have $Q(U(X_0) - U(Z)) = -Q(U(Z))$. $Q$ can be scaled such that $Q(1/k) = Q(U(X_{j+1}) - U(X_j)) = 1$ due to the fact that it is only unique up to scale. Equation (3.4) can thus be simplified to

$$Q(U(Z)) = \frac{p}{1-p}. \tag{3.5}$$

Varying $p$ then allows us to determine any number of points along $Q$. Once again, we follow BCD in choosing $p \in \{1/4, 2/5, 3/5, 3/4\}$.[9] If the elicited values of $Z$ do not happen to coincide with one of the previously elicited values from the standard sequence, $U(Z)$ is unknown. It can however be approximated through linear interpolation from the "neighboring" elements of the standard sequence in the following way:

$$U(Z) \approx U(\underline{Z}) + (U(\bar{Z}) - U(\underline{Z}))\frac{Z - \underline{Z}}{\bar{Z} - \underline{Z}} = U(\underline{Z}) + 0.2\,\frac{Z - \underline{Z}}{\bar{Z} - \underline{Z}} \tag{3.6}$$

where $\underline{Z} = max_{0 \le j \le 5}\{X_j : X_j \le Z\}$ and $\bar{Z} = min_{0 \le j \le 5}\{X_j : X_j > Z\}$.[10]

[8] Using $X_0$ or $X_1$ would typically lead to $Z$ being much larger than $X_j$, especially for large values of $p$, which would create logical inconsistencies with the framing of the choice task in the health scenario. We identified this problem in the pilot sessions and decided to restrict $X_j$ to the aforementioned, larger values.

[9] On top of the four values thus obtained, we get $Q(0) = 0$ by definition of $Q$ and $Q(0.2) = 1$ by the standardization we applied above. We thus have six values of $Q$, despite only eliciting four values of $Z$.

[10] If, for example, $X_2 < Z < X_3$, then $\underline{Z} = X_2$ and $\bar{Z} = X_3$. By construction of the lottery choice task, we must have $Z > X_0$ if $U(\cdot)$ is increasing, so that $\underline{Z}$ is always defined. However, if the regret function of an individual is strongly concave, then we may observe $Z > X_5$ so that $\bar{Z}$ is not defined. In this case we cannot interpolate but rather have to extrapolate the slope of the segment from $X_4$ to $X_5$: $U(Z) \approx U(X_5) + 0.2(Z - X_5)/(X_5 - X_4)$. In practice, this extrapolation had to be applied to around 18% of subjects.

# 4 Survey design

## 4.1 Preliminaries and pilot surveys

As previously outlined, our methodology borrows heavily from BCD. However, unlike BCD, who ran a standard, in-person lab experiment, we implemented the experiment as an online survey on the platform Limesurvey. This procedure had various practical advantages. It was possible to implement the experiment while pandemic-related restrictions rendered lab experiments infeasible. In addition, a comparatively large subject pool could be recruited from a given budget. Lastly, the online format made it easier to reach a more diverse and arguably more representative subject pool as compared to the homogeneous groups of economics students that often constitute the subject sample in economic experiments.

The subject pool was recruited on Amazon Mechanical Turk (MTurk). Amazon MTurk, the practical advantages of which are highlighted by Paolacci et al. (2010), has become a commonly used platform for recruiting experimental subjects in the social sciences (see, e.g., Arechar et al., 2018; Gandullia, 2019). Compared to standard economic laboratory experiments with student subjects, the subject pool on MTurk is generally more heterogeneous and more representative of the general US public. We are, however, aware of potential platform-specific problems that could jeopardize the quality of the acquired data. In particular, participants aiming to minimize the time they spend on the survey by "clicking through" as fast as possible can be a major problem. As discussed in section 5.3, this was also a concern in the context of our study. We therefore defined certain thresholds or sanity checks ex ante that would lead to the responses of such subjects not being used while minimizing the risk of losing observations of subjects who genuinely report their true preferences. These thresholds are explained in detail in section 4.3.

Before implementing the main survey, we ran three pilots in June and July 2021. The first of these pilots was meant to calibrate the values used in lottery choice tasks, get detailed feedback on the scenario descriptions, obtain first estimates of survey completion time and streamline the survey accordingly. 47 graduate students and recent alumni from TSE participated in this unincentivized pilot.[11] Based on the feedback and results of this first pilot, we ran a second pilot with minor adjustments, which was launched on MTurk with 100 participants. The evaluation of this second pilot made it apparent that some features of the lottery choice tasks had to be made more salient to participants without a background in economics. We made some according changes, leading to the version tested in a third and last pilot - once again launched on MTurk with 101 participants - which was, for all practical purposes, identical to the final version used for the main survey.

[11] One randomly determined participant received a €20 Amazon gift voucher to reward participation, yet the pilot was not incentivized in the sense that responses had no impact on potential payoffs.

## 4.2 Main survey

Upon accepting the task on MTurk, subjects were directed to the welcome screen on Limesurvey. This welcome screen briefly introduced them to the context and objectives of the study, indicated the expected completion time and highlighted the fact that data would be handled anonymously and that it was possible to quit the survey at any time. By continuing the survey, subjects explicitly agreed to the use of their anonymous response data. The precise wording of the welcome screen is provided in Appendix A.

After giving their consent, subjects were told which type of risks they were going to face during the survey. This could either be financial or health risks, depending on the treatment group they were assigned to. Subjects were assigned to a health scenario (H), an unincentivized financial scenario (U), or an incentivized financial scenario (I) with equal probabilities.[12] They then had to answer a small set of warm-up questions related to either financial or health risks, depending on their treatment group. An independent randomization determined the order in which subjects faced the main part of the survey: Half of all subjects started with the lottery choice tasks, i.e., the economic utility and regret elicitation, and then had to answer two short psychological regret questionnaires adopted from Schwartz et al. (2002) and Guiso (2015) (see Appendix B). For the other half, the order was inverted. This randomization was meant to control for potential priming effects of either regret elicitation method on the other. In practice, no such priming effects were observed.

The lottery choice questions began with a scenario description that differed between the three treatment groups. Subjects in the health scenario were asked to imagine that they had been diagnosed with a rare medical condition, while the financial scenario introduced a setting of lotteries with payoffs depending on randomly generated numbers. At this point, subjects in scenario I were informed that for 10% of them, one of the lottery choices they were about to make throughout the survey would randomly be chosen to determine their additional payoff. All scenario descriptions are reproduced in Appendix A. Following this, a practice question with subsequent feedback helped subjects familiarize with the structure of the lottery choice questions. After the practice question, subjects were presented a different lottery choice task and were asked to answer three comprehension questions. After attempting to answer them for the first time, they received feedback on their responses and were told to reconsider their answers for questions they had answered incorrectly. After another iteration (if necessary) of the comprehension questions and the subsequent feedback, subjects were told that they could now progress to the main part of the survey. For those in scenario I, it was highlighted again that every decision they made from now on could influence their additional payoff.

The main part consisted of a series of lottery choice questions which were interspersed with demographic questions at five predefined points. During the pilot studies, this

[12] In scenario H, the lottery choice was framed as the choice between different medical treatment options as detailed Appendix A. For simplicity, we refer to all of these as lottery choice questions in the following.

procedure had turned out to help subjects stay focused and pay closer attention to the features of each lottery choice. In particular, changes in probabilities of the respective outcomes in the regret elicitation were always preceded by an intermittent question - in addition to visual markers drawing attention to the changing probabilities, as shown in Appendix A. Around halfway through the lottery choice questions, subjects had to take the "cognitive reflection test" (Frederick, 2005). After completing all lottery choice questions, subjects who had not yet done so had to answer the psychological regret questionnaires. Lastly, subjects were given the opportunity to provide feedback through a couple of direct questions and an open feedback form.

Conditional on completing the survey, the average subject faced about 35 lottery choice questions. The precise number of questions could differ based on the choices made. The values used in each question changed dynamically based on the previous responses of the subject at hand. The details of the underlying procedure, as well as the thresholds defined on the maximum step size between elicited values, are detailed in the following section.

## 4.3 Lottery choices, thresholds and sanity checks

The measurement method explained in section 3 amounts to finding indifference values for specific lotteries. In the utility elicitation, we aim to determine $X_{j+1}$ such that $X_{{j+1}_p}g \sim X_{j_p}G$ for $j \in \{1, 2, 3, 4\}$, while in the regret elicitation, we try to find $Z_i$ such that $X_{{j+1}_{p_i}}X_0 \sim X_{j_{p_i}}Z_i$ for $p_i \in \{1/4, 2/5, 3/5, 3/4\}, X_j \in \{X_2, X_3, X_4\}$. Instead of directly asking for these indifference values, we used a choice-based procedure to determine them.[13] BCD call this procedure the "bisection method". We use essentially the same method and refer the interested reader to the original paper.

For the utility elicitation, we used $X_0 = 16$, $G = 12$, $g = 10$ and $p = 1/2$. For a given $X_j$, the elicitation of $X_{j+1}$ then started with a choice between lotteries $A = X_{j_{1/2}}12$ and $B = (X_j + 4)_{1/2}10$, i.e., the first "proposed value" of $X_{j+1}$ was $X_j + 4$. If a subject chose option A, the proposed value was increased by an amount of 4, thereby making option B more attractive. This was reiterated until they chose option B for the first time. At this point, the proposed value was reduced by an amount of 2. If a subject then chose option A (B), the proposed value was increased (decreased) by an amount of 1, leading to the last lottery choice in the elicitation of $X_{j+1}$. $X_{j+1}$ was then determined as the midpoint between the lowest value for which a subject had chosen option B and the highest value for which they had chosen option A. Table 1 illustrates the process for the elicitation of $X_1$.

If a subject always chooses option A in the above procedure, they would in theory have to face an arbitrary number of lottery choice questions and could force the proposed value of $X_{j+1}$ to become arbitrarily large. We therefore defined a threshold value of 16 on the maximum allowed step size between $X_j$ and $X_{j+1}$. If the proposed value in the elicitation reached $X_j + 16$, subjects were asked to confirm whether they indeed preferred $A = X_{j_{1/2}}12$

[13] See Bleichrodt et al. (2010) for a brief motivation of the choice-based procedure.

Table 1. Exemplary elicitation of $X_1$

| Iteration | Proposed value | Choice |
|---|---|---|
| 1 | 20 | B |
| 2 | 18 | A |
| 3 | 19 | A |
| $X_1$ | 19.5 | |

over $B = (X_j + 16)_{1/2}10$. We refer to such behavior as "unreasonable risk aversion" (URA). If they confirmed, they "dropped out" of the survey in the sense that they faced no further lottery choice questions and their responses were not used in the analysis of utility and economic regret. Similarly, if a subject always chooses option B in the above procedure, they face the choice between $A = X_{j_{1/2}}12$ and $B = X_{j_{1/2}}10$ after three consecutive choices of B. If they confirmed a preference for option B, they were equally removed from the elicitation. This behavior is labeled "violation of statewise dominance" (VOSD) in the following.

The threshold value of 16 that was used to determine URA was not chosen arbitrarily, but rather determined based on data from the pilot studies. VOSD clearly cannot reflect true preferences and must thus be a random decision mistake. In the context of our survey, URA is symmetric to VOSD in the sense that it requires subjects to always choose the same option. If URA is a random decision mistake as well, instead of reflecting true preferences, we would expect to see similar shares of subjects being removed by either criterion. We thus defined a threshold value such that the number of subjects violating either criterion was approximately equal in the pilot sessions. Based on the student pilot, this threshold value on the step size was 8. However, subjects on MTurk generally had larger step sizes so that we decided to double the maximum allowed step size to 16 after the second pilot. With this threshold, we are confident that we do not remove subjects who genuinely report their true preferences.

The elicitation of the regret function was designed in an almost identical way. For $p_i \in \{1/4, 2/5\}$, the elicitation of $Z_i$ started with a choice between $A = X_{j+1_{p_i}}X_0$ and $B = X_{j_{p_i}}(X_0 + 4)$ for given $X_j$ and $X_{j+1}$ with $X_j \in \{X_2, X_3, X_4\}$ randomly determined. The first proposed value for $Z_i$ was $(X_0 + 4)$ and then progressed just as for utility, with the proposed values of $Z_i$ changing accordingly. For $p_i \in \{3/5, 3/4\}$, the first proposed value for $Z_i$ was $(X_0 + 8)$ instead, and the initial step size was 8 instead of 4. The only other difference compared to utility elicitation is that we made regret elicitation "more precise" by a factor of 2. The elicitation did not stop when the difference between the lowest value for which a subject had chosen option B and the highest value for which they had chosen option A was 1, but only when it was 0.5. Table 2 clarifies the procedure with an exemplary elicitation of $Z_4$.

Table 2. Exemplary elicitation of $Z_4$

| Iteration | Proposed value | Choice |
|---|---|---|
| 1 | 24 | A |
| 2 | 32 | B |
| 3 | 28 | B |
| 4 | 26 | A |
| 5 | 27 | A |
| 6 | 27.5 | A |
| $Z_4$ | 27.75 | |

*Table notes*: in this example, $p = 3/4, X_j = 21, X_{j+1} = 24.5$.

As a direct test of the predictions of RT, we added a single lottery choice question at the end the utility elicitation which we refer to as a consistency check. RT predicts that subjects must be indifferent between $A = X_{4_{1/2}}X_0$ and $B = X_{3_{1/2}}X_1$ since $Q(U(X_4) - U(X_3))/2 = Q(U(X_1) - U(X_0))/2$ where $U(X_4) - U(X_3) = U(X_1) - U(X_0)$ by construction of the standard sequence. If we make lottery B more attractive by increasing one of its payoffs - sufficiently so to allow for some margin of error - then a subject behaving in line with RT must prefer lottery B over lottery A. We achieved this by replacing $X_1$ in lottery B with $X_1 + 2 + (X_4 - X_3)/4$, rounded to the nearest half integer. We thus allowed for a margin of error of at least 2, plus a variable part accounting for potentially larger errors as the differences between payoffs increase. In practice, the lottery choice was between $A = 27_{1/2}16$ and $B = 24.5_{1/2}21$ for the median subject.

# 5 Subjects, dropouts and response time

## 5.1 Subject sample

A total of 1004 subjects completed the survey. Among them, 14 completed the survey twice and were excluded from the analysis. Another 4 subjects took more than one hour to complete the survey and were also excluded, leaving 986 observations (63% male). 311 subjects faced scenario H, while 339 (336) subjects faced scenario U (I). Mean and median response times were just above 20 minutes, perfectly in line with the estimated response time of 15 to 25 minutes that we indicated to subjects on the welcome screen.

The mean age of subjects was 38 years, with 83% between 25 and 50 years old. 71% of subjects indicated they were married or in a domestic partnership. 85% indicated they held at least a bachelor's degree. More than 80% of subjects were in full-time employment, with the median reported annual gross household income between \$50,000 and \$70,000. While we cannot claim to have a perfectly representative sample of the adult US population, our subject pool is certainly more diverse and representative than the typical sample of undergraduate (economics) students that are frequently employed in economic lab experiments.

## 5.2 Payments

Subjects received a payment of $2 for their participation in the experiment. While this sum may seem quite small compared to show-up fees in standard laboratory experiments, there are multiple reasons to believe that it was nonetheless sufficient to motivate subjects. First of all, hourly wages in the range of $5 to $10 are very common on Amazon MTurk (Auer et al., 2021). Secondly, the target of 1,000 responses was almost reached within two hours of setting the task online, indicating that subjects were eager to take part in the experiment based on the expectation of earning $2 in 15 to 25 minutes. Thirdly, only six subjects indicated that they were dissatisfied with the payment relative to the time they spent.[14] We therefore argue that that the relatively low participation fee did not systematically lead to subjects paying no attention or showing "protest behavior", while allowing us to generate many observations from a given budget.

Additional payments were made to subjects in the incentivized financial treatment. After facing the welcome screen and answering the "financial warm-up questions", they were told that for 10% of all subjects, one randomly selected lottery would be implemented and paid out based on their choices and a randomly generated number. This procedure is similar to the one used by BCD, except that we only applied it to one out of three treatment groups.[15] 37 subjects received additional payments, ranging from $10 to $48 (mean = $18.07).

## 5.3 Dropouts

Before subjects faced the lottery choice questions that we analyze here, we tested their understanding of the task at hand and the resulting payoffs. The three "comprehension questions" we used can be found in Figure A4.1 in Appendix A. After their first attempt at answering these questions, subjects were given another chance to change their answer for questions they had previously answered incorrectly, as detailed in Figure A4.2. If even after the second attempt they did not answer the comprehension questions correctly, we did not include their data in our analyses. This led us to remove 367 subjects, 148 of whom also "dropped out" of the experiment due to unreasonable answers as detailed below. As explained in section 4.3, there were two potential reasons for subjects to "drop out" of the experiment: Choosing a lottery that is statewise dominated by the alternative (VOSD) and unreasonable risk aversion (URA). Subjects who dropped out for one of these reasons faced no further lottery choice questions; however, they still had to answer all other remaining

[14] This is despite the fact that more than two thirds of subjects left open text feedback with an average length of 26 characters. It is therefore reasonable to assume that subjects would have been more outspoken about the issue if they had indeed perceived the payment as insufficient or unfair.

[15] For a general discussion of the validity of this incentive system, see Cubitt et al. (1998).

questions (e.g., demographic and follow-up questions), depending on where exactly they dropped out. Observations of subjects who did not complete all blocks of both the utility and regret elicitation were dropped in the analysis of both utility and regret functions. Importantly, subjects were not told they had dropped out or done anything "wrong", and still received the standard payment as well as - where applicable - the additional payment in the incentivized financial treatment.

Throughout the experiment, a total of 312 subjects dropped out - 217 (95) of which due to URA (VOSD) - leaving 674 complete observations. An additional 219 subjects were removed because of incorrect responses to the comprehension questions even after two attempts. Dropout rates across the three scenarios are displayed in Table 3. Total dropout rates across scenarios are almost identical, and neither pairwise t-tests nor a chi-squared test reveal any significant differences ($p > 0.10$). Almost 40% of the dropouts occurred in the first block alone, that is, in the elicitation of $X_1$.[16] In practice, dropping out in the first block is equivalent to one of the following two preferences: $16_{0.5}10 \succeq 16_{0.5}12$ (VOSD) or $16_{0.5}12 \succeq 32_{0.5}10$ (URA), with the numbers referring either to dollar amounts or life expectancy in months. At least for monetary outcomes, it is clear that neither of these should reflect true preferences - they are thus treated as decision mistakes. Since the share of subjects dropping out is not significantly different for subjects facing health outcomes, we conclude that dropouts there are similarly driven by mistakes.

In order to drop out, a subject had to choose the same option - either A or B - at least four consecutive times and was asked to confirm their choice on the last question that led to their dropping out. We therefore suspect that a large share of dropouts are due to subjects simply clicking through the lottery choice questions as fast as possible, always selecting the same answer. This conjecture can be backed up by a closer investigation of the relationship between response time and dropouts.

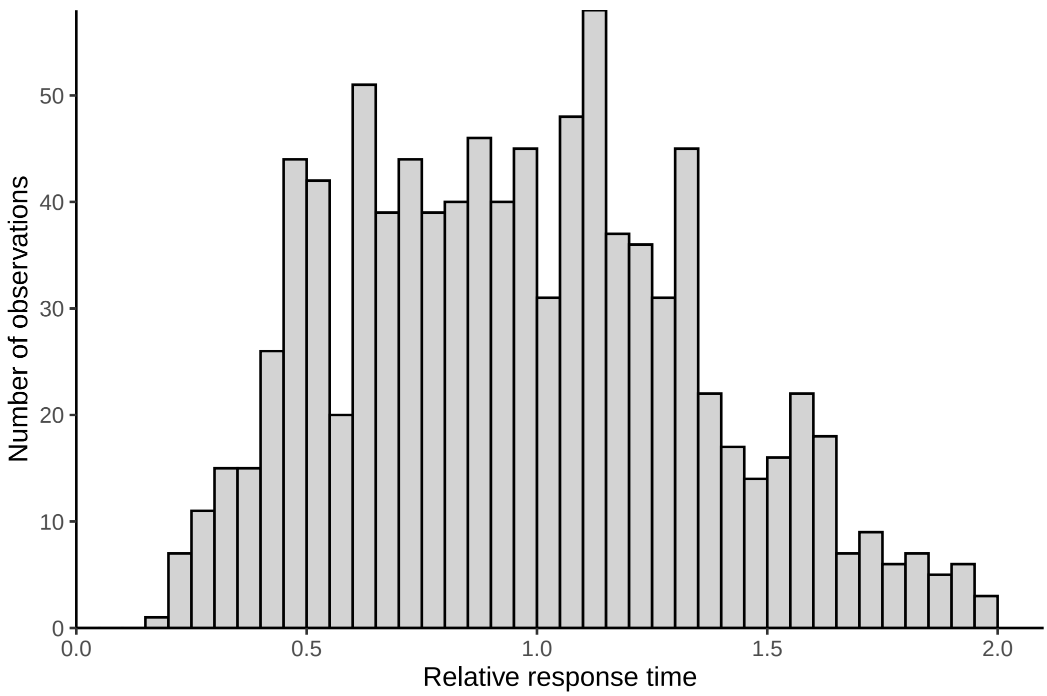
Since subjects that dropped out did not have to answer any of the remaining lottery choice questions, their total response time is automatically lower compared to subjects who completed the entire survey. The relationship between response time and dropout rates is therefore bidirectional, and naively regressing the probability to drop out on total response time would not yield valid results. To overcome this issue, we use the relative response

Table 3. Number of subjects that dropped out of the elicitation based on VOSD and URA by scenario

| Scenario | Total subjects | URA | VOSD | Incomprehension | Total dropouts | Remaining |
|---|---|---|---|---|---|---|
| H | 311 | 80 | 29 | 65 | 174 | 137 |
| U | 339 | 65 | 36 | 78 | 179 | 160 |
| I | 336 | 72 | 30 | 76 | 178 | 158 |
| Total | 986 | 217 | 95 | 219 | 531 | 455 |

[16] In subsequent blocks starting from the third - after the elicitation of $X_2$ - dropout rates were fairly low and homogeneous, ranging between 1.7% and 3.4% from one block to another.

Figure 1. Histogram of relative response time




time instead, defined as the quotient of total response time and the mean response time for all questions that subject $i$ actually faced:

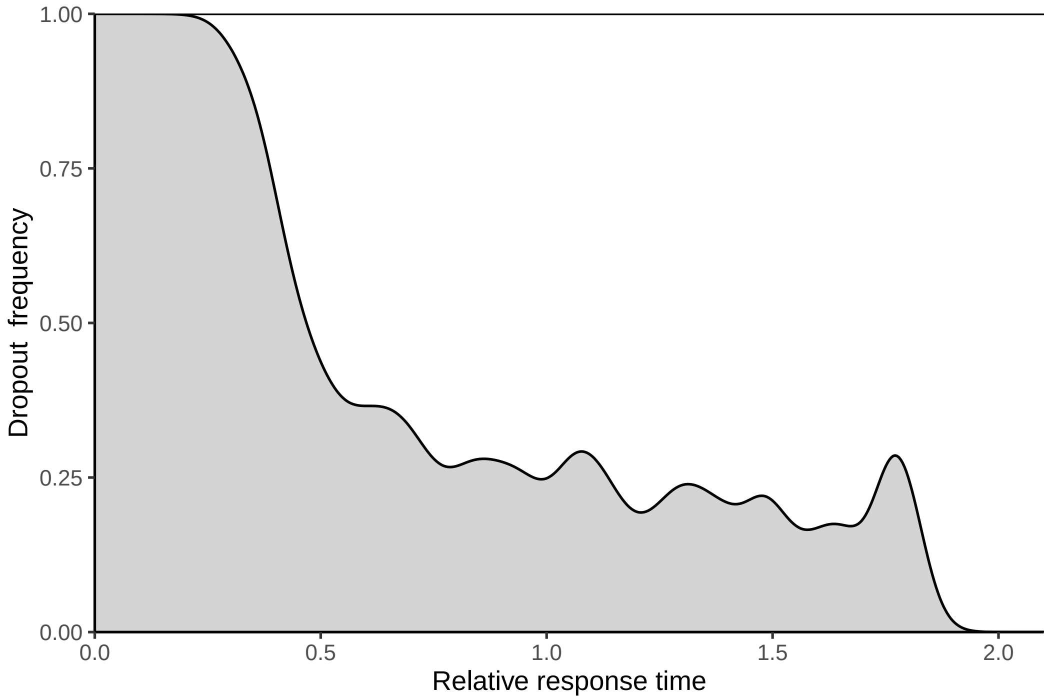
$$\hat{t}_i = \frac{T_i}{\sum_{q \in Q_i} \bar{t}_q} \tag{5.1}$$

where $T_i$ is the total response time of subject $i$, $Q_i$ is the set of questions faced by subject $i$ and $\bar{t}_q$ is the mean response time for question $q$ across all subjects facing this question. A value of 1 would indicate that subject $i$ took exactly the average amount of time in responding to the questions she faced.[17] A substantially lower (higher) value indicates that the subject was substantially faster (slower). Figure 1 shows the distribution of relative response times, while Figure 2 shows a kernel density estimate of the respective dropout rates as a function of response time. 23 observations with a relative response time larger than 2 are not displayed in the two graphs.

Figure 2 illustrates a strong negative relationship between relative response time and dropout rates for relative response times substantially below 1. For average and above average response times, the relationship is less clear. During our pilot sessions, we had identified a total response time of around 10 to 15 minutes as the bare minimum necessary for an experienced survey participant to read and answer all questions, without spending any additional time pondering on what to answer. This corresponds to a relative response

[17] For a subject who completed the survey without dropping out, this average would typically correspond to a total response time of around 22 to 23 minutes. Some variation remains due to the fact that even among subjects who did not drop out, the number of lottery choice questions they faced was not fixed but rather depended on their answers, as described in section 4.3.

Figure 2. Kernel density estimate (gaussian) of dropout rates as a function of relative response time




time of around 0.5. Subjects with a lower relative response time may not have read all of the instructions and questions, and most certainly did not spend any time reflecting on their answers. This is reflected in the fact that 71.4% of subjects with a relative response time of less than 0.5 dropped out, while only 26.2% of those taking more time dropped out. Taking a slightly more conservative threshold of 0.67, we can explain 130 out of 312 (41.7%) dropouts based on low response time alone.[18] In short, almost half of the dropouts can be attributed to subjects not taking an adequate amount of time to answer the questions but instead rapidly clicking through the survey in order to claim the participation fee without exerting effort.

# 6 Results

In this section, we first describe the measures of curvature used to classify subjects' utility and regret functions as concave, linear or convex, and then present aggregate and individual-level results of utility elicitation for the three different scenarios, denoted by H (health outcomes), U (uninventivized financial outcomes) and I (incentivized financial outcomes). Finally, we turn to economic regret aversion, while testing for differences between health and financial outcomes. Unless stated otherwise, differences between two

[18] A relative response time of 0.67 corresponds to around 15 minutes for a subject completing the survey without dropping out. This is the lower boundary we indicated in the instructions and the minimum amount of time we deem necessary to carefully read the scenario descriptions and think about answers for at least a short moment.

mutually exclusive and exhaustive proportions from the same sample are tested with the binomial test, while differences between non-exhaustive proportions from the same sample are tested using a normal approximation (see Scott and Seber, 1983).

## 6.1 Measures of curvature

For both utility and regret functions, we closely follow BCD and apply three different measures of curvature.[19] We obtained parametric estimates of utility curvature by fitting a power function of the form $U = \alpha(X - 16)^\beta$ to each subject's elicited values by nonlinear least squares (NLS) estimation. The utility function U was scaled to $U(16) = 0$ and $U(X_5) = 1$ for each subject, thereby normalizing its range to $[0, 1]$. The second measure is the area under the standardized utility curve, once again scaling the range of $U$ to $[0, 1]$. The domain was similarly rescaled to $[0, 1]$ by transforming the $X_i$ to $\hat{X}_i = (X_i - 16)/(X_5 - 16)$ for each subject. An area larger than 0.5 corresponds to locally concave utility, while an area below 0.5 indicates locally convex utility. A linear utility function yields an area of exactly 0.5. The third measure of utility curvature is based on pairwise comparisons of the slope of different segments of the utility function. We define

$$\begin{gathered} \Delta_{gh,lm} = (X_g - X_h) - (X_l - X_m), \\ \forall g, h, l, m \in \{0, ..., 5\} \text{ with } g > h,\ g > l,\ g - h = l - m. \end{gathered} \tag{6.1}$$

This yields a total of 20 values for each subject. Since $g - h = l - m$, $U(X_g) - U(X_h) = U(X_l) - U(X_m)$ by definition of the standard sequence. A linear utility function should therefore yield a value of 0 for each $\Delta_{gh,lm}$. Positive (negative) values indicate concavity (convexity) of the utility function since for a concave (convex) function, larger (smaller) consecutive increases of the argument are needed to achieve a given increase in utility. A subject's utility function is therefore classified as concave (convex) if more than 10 values of $\Delta_{gh,lm}$ are larger (smaller) than zero. If more than 10 values of $\Delta_{gh,lm}$ are equal to zero, the utility function is classified as linear, and otherwise it is left unclassified.

Similar measures of curvature are used to classify subjects' regret functions. To obtain parametric estimates of regret curvature, we attempted to fit a power function of the form $Q = \alpha U^\beta$ for each subject using NLS. Unfortunately, this estimation did not converge for many subjects, and estimated coefficients were sensible to the chosen start values. We therefore decided not to use this measure in the following analysis. To obtain the area under the standardized regret curve, we rescaled the domain to $[0, 1]$ by transforming

[19] In fact, all our measures of curvature are identical to the ones used by BCD, except for the pairwise comparison of slopes for the regret function. For the latter, BCD do not provide a valid definition of their measure, since they claim to use the values $Q(i/5), i \in \{0, ...5\}$. This may be an artifact from the probability elicitation procedure they used in pilot sessions and earlier versions of their paper, yet it is not applicable when $Q$ is determined with the outcome elicitation procedure since in general, no values for $Q(2/5)$, $Q(3/5)$, $Q(4/5)$ and $Q(1)$ are available.

$U(Z_i)$ to $\hat{U}(Z_i) = U(Z_i)/max\{U(Z_j)\}$ with $j \in \{0, ..., 5\}$ for each subject.[20] For the third and last measure of regret curvature, we once again make pairwise comparisons of slope, taking into account the fact that the $U(Z_i)$ are not evenly spaced in general. We define

$$\nabla_{gh,lm} = \frac{Q(U(Z_g)) - Q(U(Z_h))}{U(Z_g) - U(Z_h)} - \frac{Q(U(Z_l)) - Q(U(Z_m))}{U(Z_l) - U(Z_m)}, \quad \forall g, h, l, m \in \{0, ..., 5\} \text{with } g > h,\ g > l,\ g - h = l - m. \tag{6.2}$$

Again, this yields 20 values per subject. Along a convex (concave) regret function, the slope is increasing (decreasing) so that positive (negative) values of $\nabla_{gh,lm}$ indicate convexity (concavity). Due to the nature of the elicitation mechanism, which only determines approximate indifference values, $\nabla_{gh,lm}$ will in general not be 0 even if the true regret function of a subject is approximately linear. We therefore cannot classify subjects' regret functions as linear based on this measure. Furthermore, a linear regret function would in expectation yield an equal number of negative and positive values of $\nabla_{gh,lm}$, so that a threshold of 50% of values being positive to classify a regret function as convex is clearly inadequate. We instead classify a regret function as convex (concave) if at least 14 out of 20 values are positive (negative), but also provide results for different cutoffs.[21]

## 6.2 Results on utility elicitation

531 subjects were removed due to URA, VOSD or incorrect responses to the comprehension questions and are therefore not included in the following analyses, leaving 455 observations (137 H, 160 U, 158 I). Graphs of the elicited utility functions for each scenario using median or mean data are provided in Appendix C. Median utility functions in scenarios H and I are perfectly linear. In scenario U, the graph does not exhibit a clear pattern and is again close to linear. When using mean data, utility in scenarios U and I appears to have a slightly more convex shape, while utility in scenario H remains roughly linear. This graphical intuition can be supported by fitting a simple power function to the respective mean or median observations of the $X_i$ for each scenario using NLS. The results of this estimation are reported in Table 4.

The estimated curvature based on mean data indicates relatively low but significant convexity of utility in the two financial scenarios. In scenario H, curvature is not significantly different from 1. Using median data, we find linear utility for all three scenarios. With mean data, utility curvature is significantly lower for H than for U ($p < 0.01$), and for I (p

[20] In theory, we expect $max\{U(Z_j)\} = U(Z_4)$ since the regret function should be strictly increasing, regardless of its being convex, linear or concave. This is indeed the case for the majority of subjects. However, since we allow for some response error, we observe $max\{U(Z_j)\} > U(Z_4)$ for some subjects. Notice that $U(Z_0) = 0$ with $Q(U(Z_0)) = 0$ and $U(Z_5) = 0.2$ with $Q(U(Z_5)) = 1$ due to the scaling of Q explained in section 3.

[21] Using lower thresholds leads to a larger number of subjects being classified at all, thereby yielding higher statistical power and more significant results when comparing the respective proportions. We take a conservative approach that leaves a relatively large share of subjects unclassified.

Table 4. Parametric estimates of mean and median utility functions

| Scenario | Mean data | Median data |
|---|---|---|
| | $\hat{\beta}$ | $\hat{\beta}$ |
| H | 1.0504* | 1′ |
| | (0.0219) | (0) |
| U | 1.1922*** | 1.0721 |
| | (0.0226) | (0.0546) |
| I | 1.1195*** | 1′ |
| | (0.0206) | (0) |

*Table notes*: Standard errors in parentheses. Significance levels are reported for testing the null hypothesis that $\hat{\beta} = 1$. The notations for significance levels are as follows: * for p≤0.1; ** for p≤0.05; *** for p≤0.01; ′ for perfect fit through the six elicited points.

= 0.02). With median data, there are no significant differences between the curvatures. Overall, the aggregate data indicate that subjects tend to show slightly stronger risk aversion or less risk seeking for health risks than for financial risks, yet the differences are small in magnitude.

Strategic responding - a potential concern in the incentivized scenario - did not play any noticeable role. If subjects in scenario I had tried to increase their potential payoffs by choosing the "safer" lottery even if it did not reflect their true preferences - which would lead to their being offered a more attractive "risky" lottery in subsequent choices - we would expect to see larger values of the respective $X_i$ in scenario I as compared to scenario U. However, pairwise comparisons of all $X_i$ across scenarios do not reveal any significant differences in the respective means and distributions, using t-tests and Kolmogorov-Smirnov tests.

Tables 5 shows the classification of individual utility functions based on the three measures of curvature discussed above. In scenario H, there is no significant difference between the share of subjects showing convex and concave utility using any of the three measures ($p > 0.1$ for all three measures). In scenario U, the share of convex subjects is significantly higher than the share of concave subjects for all measures of curvature ($p < 0.01$ each). The same holds true for scenario I, with the exception of the parametric estimates where the shares do not differ significantly.

Pairwise comparisons between scenarios reveal that the share of concave subjects is consistently highest in scenario H and lowest in scenario U. The difference in shares between scenarios H and U is significant using the parametric and slope-based measures of curvature ($p < 0.05$ each). The latter also indicates a significant difference between scenarios H and I ($p = 0.025$). None of the differences between scenarios U and I is significant. If anything, subjects in the unincentivized scenario showed more risk averse

Table 5. Classification of individual utility functions

(a) based on parametric estimation

| | Convex | | Concave | Linear | Unclassified | Total |
|---|---|---|---|---|---|---|
| H | 24 | $\overset{*}{\longleftrightarrow}$ | 39 | 28 | 46 | 137 |
| U | 58 | $\overset{***}{\longleftrightarrow}$ | 31 | 20 | 51 | 160 |
| I | 29 | | 32 | 38 | 59 | 158 |
| Total | 111 | | 102 | 86 | 156 | 455 |

(b) based on the area under the curve

| | Convex | | Concave | Linear | Total |
|---|---|---|---|---|---|
| H | 49 | | 52 | 36 | 137 |
| U | 86 | $\overset{***}{\longleftrightarrow}$ | 50 | 24 | 160 |
| I | 69 | $\overset{**}{\longleftrightarrow}$ | 42 | 47 | 158 |
| Total | 204 | $\overset{***}{\longleftrightarrow}$ | 144 | 107 | 455 |

(c) based on the slope evolution criterion

| | Convex | | Concave | Linear | Unclassified | Total |
|---|---|---|---|---|---|---|
| H | 33 | | 38 | 31 | 35 | 137 |
| U | 61 | $\overset{***}{\longleftrightarrow}$ | 32 | 20 | 47 | 160 |
| I | 45 | $\overset{**}{\longleftrightarrow}$ | 28 | 42 | 43 | 158 |
| Total | 139 | $\overset{***}{\longleftrightarrow}$ | 98 | 93 | 125 | 455 |

*Table notes:* Significance levels are reported for testing the null hypothesis that the share of convex and concave subjects coincide. The notations for significance levels are as follows: * for p≤0.1; ** for p≤0.05; *** for p≤0.01.

behavior - the opposite of what we would expect if strategic responding played a role in the incentivized scenario.

Overall, the three measures of utility curvature provide coherent results, despite some noise in the data. The notion that on average, preferences for health outcomes are more concave or less convex than for financial outcomes is supported by the individual-level data.

## 6.3 Results on regret attitude

Following BCD, we demand that subjects satisfy monotonicity. Monotonicity requires that if $p > q$ and $X_{i_p}X_0 \sim X_{i-1_p}Z$ and $X_{i_q}X_0 \sim X_{i-1_q}Z'$, then $Z > Z'$. Intuitively, as the probability of receiving the better outcome $X_i$ increases, the alternative act must become more attractive through an increase in $Z$ to maintain indifference. This implies that the elicited values of $Z_i$ should be increasing. This is also a requirement of RT in the sense that the regret function, regardless of its precise shape, must be increasing.

Table 7. Parametric estimates of mean and median regret functions

| Scenario | Mean data | Median data |
|---|---|---|
| | $\hat{\beta}$ | $\hat{\beta}$ |
| H | 1.1129 | 1.2559*** |
| | (0.0991) | (0.0479) |
| U | 0.9411 | 1.5466** |
| | (0.0639) | (0.1669) |
| I | 1.0847 | 1.3956** |
| | (0.0981) | (0.1273) |

*Table notes*: Standard errors in parentheses. Significance levels are reported for testing the null hypothesis that $\hat{\beta} = 1$. The notations for significance levels are as follows: * for p≤0.1; ** for p≤0.05; *** for p≤0.01.

Like BCD, we allow for a margin of error of half the average step size in the elicitation of the first two values of the standard sequence, $(X_1 + X_2)/2$. Discarding observations that do not satisfy this criterion leaves us with 312 observations (85 H, 107 U, 120 I). Violations of monotonicity are strongly correlated with incorrect responses to the comprehension questions. Among subjects who answered all comprehension questions correctly, only 31.4% violated monotonicity, while 61.2% of those who did not answer correctly even after two attempts violated monotonicity. By applying the monotonicity criterion, we therefore primarily remove subjects who did not fully understand the task or did not think seriously about their answers. In Appendix E, we report robustness checks for both an unrestricted sample and an even more restrictive cutoff rule, allowing for no margin of error.

As with utility, it is instructive to begin the analysis of regret on an aggregate level. Plots of the mean and median regret functions are provided in Appendix D. Most of the elicited mean values lie below a straight line connecting $Q(0) = 0$ and $Q(U(Z_4)) = 3$ and could therefore be classified as convex, using the term rather leniently. It is, however, apparent that we do not obtain such well-behaved, clearly convex functions as BCD did. Median data provide stronger support for convexity. As for the average utility functions, we fitted a power function of the form $Q = \alpha U(Z)^{\beta}$ to the respective mean and median observations. Results for both types of aggregate data and each scenario are reported in Table 7.

None of the estimates based on mean data are significantly different from 1. Using median values instead leads to significantly convex regret function estimates for all three scenarios, with the most precisely estimated coefficient in scenario H. None of the estimated curvatures differ significantly in pairwise comparisons. Obviously, these results based on aggregate data are purely indicative and need to be backed up by individual-level analyses.

In fact, a more disaggregated analysis reveals that there is no clear support for convex regret functions.

Tables 9a and 9b show the classification of individual regret functions based on both measures of curvature discussed above. As expected, both measures are strongly correlated ($\rho = 0.624$).[22] Based on the area under the curve, the share of convex subjects is significantly higher than the share of concave subjects when pooling across scenarios ($p < 0.01$), with the only significant individual difference in scenario I. None of the shares differ significantly from each other across scenarios. Based on slope evolution, however, the majority of subjects are rated concave in each of the scenarios. Again, the proportion of concave and convex subjects only differs significantly in scenario I, now in the opposite direction. When aggregating across scenarios, the difference is again highly significant ($p < 0.01$). Applying different thresholds for subjects to be classified as convex or concave substantiates this notion, as detailed in Table 9c. Requiring any number between 11 and 16 out of 20 values of $\nabla_{gh,lm}$ to be positive (negative) for a subject to be classified as convex (concave), we find a significantly higher share of concave subjects when pooling across scenarios up to a cutoff of 14 ($p < 0.01$ each). For even larger cutoff values, most subjects cannot be classified and the classification is thus underpowered.

Taken together, the two measures of regret attitude do not yield conclusive evidence regarding the predominance of convex or concave regret functions. As is well-known, RT requires the regret function to be convex for sensible behavioral predictions. Even when giving the benefit of the doubt to the area under the curve - the measure reporting more convexity - around 42% of subjects exhibit concave regret functions. When relying on the slope evolution criterion, at least 70% of subjects do not have convex regret functions, depending on the precise cutoff value chosen. A very large share of subjects therefore do not behave in accordance with RT. This share becomes even larger when incorporating the "consistency question" introduced in section 4.2. The latter was constructed so that, based on the predictions of RT, subjects should have a clear preference for one of the options even after allowing for a substantial margin of error. Out of 312 subjects for whom we defined regret measures in this section, 66 violate the predictions of RT in this lottery choice. Requiring that subjects answer this consistency question as RT predicts and that their regret function is convex as measured by the area under the curve, only 138 out of 312 (44%) subjects behave in accordance with RT. Using the slope evolution criterion, only 67 (21.5%) subjects behave in line with RT, even when applying a very lenient threshold and requiring only 11 values of $\nabla_{gh,lm}$ to be larger than 0. While some of these surprising results may be due to noise and decision mistakes instead of reflecting

[22] $\rho$ is determined as the Spearman correlation coefficient between the area under the standardized regret curve and the number of negative values of $\nabla_{gh,lm}$. For a convex (concave) function, the area under the curve should be low and there should be few negative values of $\nabla_{gh,lm}$, so that we expect a strong positive correlation. When considering a simple $2 \times 2$ rank correlation - that is, classifying subjects only as "convex" and "non-convex" based on both measures - correlation is naturally weaker but still substantial at $\rho = 0.376$.

subjects' true preferences, it is clear that RT does not describe the choices of a large share of subjects well.

Table 8. Classification of individual regret functions

(a) based on the area under the curve

| | Convex | | Concave | Total |
|---|---|---|---|---|
| H | 52 | $\overset{*}{\longleftrightarrow}$ | 33 | 85 |
| U | 55 | | 52 | 107 |
| I | 74 | $\overset{**}{\longleftrightarrow}$ | 46 | 120 |
| Total | 181 | $\overset{***}{\longleftrightarrow}$ | 131 | 312 |

(b) based on the slope evolution criterion (cutoff = 14)

| | Convex | | Concave | Unclassified | Total |
|---|---|---|---|---|---|
| H | 21 | | 27 | 37 | 85 |
| U | 14 | $\overset{*}{\longleftrightarrow}$ | 26 | 67 | 107 |
| I | 16 | $\overset{**}{\longleftrightarrow}$ | 31 | 73 | 120 |
| Total | 51 | $\overset{***}{\longleftrightarrow}$ | 84 | 177 | 312 |

(c) based on the slope evolution criterion (different cutoffs, pooled across scenarios)

| Cutoff value | Convex | | Concave | $p \leq 0.05$ in scenario |
|---|---|---|---|---|
| 11 | 94 | $\overset{***}{\longleftrightarrow}$ | 183 | U, I |
| 12 | 80 | $\overset{***}{\longleftrightarrow}$ | 150 | U, I |
| 13 | 62 | $\overset{***}{\longleftrightarrow}$ | 117 | I |
| 14 | 51 | $\overset{***}{\longleftrightarrow}$ | 84 | I |
| 15 | 35 | $\overset{**}{\longleftrightarrow}$ | 59 | I |
| 16 | 23 | | 32 | |

*Table notes:* Significance levels are reported for testing the null hypothesis that the share of convex and concave subjects coincide. The notations for significance levels are as follows: * for p≤0.1; ** for p≤0.05; *** for p≤0.01.

Table 10. Overlap of both measures of regret curvature

(a) slope cutoff = 14

| | Convex slope | Unclassified slope | Concave slope |
|---|---|---|---|
| Convex area | 51 | 112 | 18 |
| Concave area | 0 | 65 | 66 |

(b) slope cutoff = 11

| | Convex slope | Unclassified slope | Concave slope |
|---|---|---|---|
| Convex area | 87 | 23 | 71 |
| Concave area | 7 | 12 | 112 |

### 6.4 Correlates of regret attitude

Finding subgroups with particular characteristics for whom RT can accurately predict behavior would allow RT to redeem itself at least partially. Since we have no clear ex ante predictions as to what characteristics may reasonably affect regret sensitivity, this analysis is exploratory and subject to the problems of multiple hypothesis testing. Its results can be summarized very neatly: None of the demographic variables and other covariates we collected have any predictive power for regret attitudes as measured by either the area under the curve or the slope evolution.

We regressed both of the regret measures[23] on each covariate individually and on all covariates in a joint regression, always including scenario fixed effects. Since many of the covariates are ordinal factors with moderately many levels, this ultimately leads to the estimation of 45 coefficients in the joint regression. Clearly, significance levels must be corrected for based on multiple hypothesis testing methods in this context (Romano and Wolf, 2005; List et al., 2019). In particular, when setting the family-wise error rate to be $\alpha = 0.05$, the "most significant effect" must have $p < \alpha/\#\text{coefficients} \approx 0.001$ in our case. In either of the joint regressions we do not find any of the estimated coefficients to be nearly as significant. The number of "seemingly significant" coefficients is perfectly in line with what would be expected by chance alone when regressing on this many covariates with no true effect. We therefore conclude that our measures of regret sensitivity do not systematically relate to demographics and other individual-level covariates. This substantiates the notion that they may after all not measure any meaningful concept underlying the preferences of our subjects.

# 7 Discussion

Though parametric estimates using median data seem to confirm BCD's finding that individuals are regret averse on an aggregate level, our results at the individual level differ significantly from theirs. More specifically, BCD documented that convexity was the dominant pattern at the individual level.[24] We find no convincing evidence of such a pattern. Using the area under the standardized regret curve as the measure of regret sensitivity, we do find a significantly higher share of convex than concave subjects when pooling across scenarios. This finding is reversed, with the shares of subjects being more strongly skewed in favor of concave regret functions when using the slope evolution criterion.

[23] The area under the standardized regret curve could immediately be used as the dependent variable, with higher values corresponding to a more concave function. To operationalize the slope evolution criterion as a single outcome variable, we computed the number of $\nabla_{gh,lm} > 0$ minus the number of $\nabla_{gh,lm} < 0$, which could in theory take any value between -20 and 20 and again has higher values corresponding to a more concave function.

[24] Based on the slope evolution criterion, BCD found 30 (36) subjects were convex and 10 (15) were concave in their first (second) measurement. Based on the area under the standardized regret curve, they found 37 (41) subjects were convex and 4 (5) concave in their first (second) measurement. The proportion of convex subjects was always significantly higher than the proportion of concave subjects ($p < 0.01$ in all tests).

These findings remain qualitatively unchanged when using a less restricted sample - not requiring any form of monotonicity of the regret function - or a much more restricted sample - requiring montonicity without any margin of error.

One might immediately point out that this may be due to differences in conducting the experiments. Indeed, our study differs from BCD in three main ways. First, our subjects recruited on Amazon M-turk are more diverse (and thus more representative) than the student sample in BCD. Second, our experiment was conducted online but BCD's was in a physical lab. Last but not least, only subjects in scenario I faced real incentives for their choices. However, we believe that none of these differences could explain the difference between our results.

Regarding the sample pool, we think this is actually a contribution we make to the literature. If regret aversion is indeed a prominent feature of individual preferences, we should expect it to extend not only to students, but to the more general population as well. Regarding the venue of doing the experiments, MTurk participants are perhaps less attentive and noisier than their student counterparts (Fréchette et al., 2021). However, as illustrated in section 6.3, it is unlikely that our different results were driven by decision mistakes given our significant efforts to exclude inattentive subjects. Moreover, average time spent by MTurk participants and by student subjects in the pilot sessions was almost the same. This further rules out the possibility that online experimentation is the source of different results. Finally, since subjects in scenario I faced comparable incentives as in BCD,[25] it is difficult to link our different results with a lack of incentives. Besides, since we have found no significant difference in regret attitude between scenarios U and I, the existence of direct incentives on choices does not seem to really affect subjects' behavior.[26]

Since we have obtained very different findings compared to BCD, it is also important to discuss the main features of the experimental design used in both studies. First of all, both studies have only tested original regret theory (Loomes and Sugden, 1982) rather than generalized regret theory (Loomes and Sugden, 1987a). Since original regret theory has more structure (and therefore more predictive power), it is also easier to falsify than generalized regret theory. So, could our negative results on RT arise due to incorrect structural assumptions about the regret function? Recent work by Loewenfeld and Zheng (2023) has taken a non-parametric approach to directly test generalized regret theory and conducted both lab experiments with Chinese students and an online experiment with a more representative Chinese sample. The authors also found no positive evidence in favor of RT. Drawing on their findings, we believe that our negative results on RT are not attributable to inappropriate structural assumptions over preferences.

[25] In BCD, six subjects among 55 (10.9%) were randomly selected to play one of their choices for real at the end of the experiment. So the chance that a subject receives additional payments is fairly close between our study and BCD.

[26] In the literature of valuation of non-market goods such as health and environment, researchers often use contingent valuation methods with hypothetical questions to elicit individuals' willingness to pay. Many studies have found no disparity between real and hypothetical willingness to pay (e.g., Carlsson and Martinsson, 2001), namely, respondents truthfully report their preferences regardless of incentives.

Another key experimental feature is that during our experiment and also BCD's, subjects received no feedback on their choice outcome and what could have been obtained had they chosen differently. They were only informed of their final earnings at the end of the experiment. In the psychology literature, it is often argued that in order for people to anticipate regret and act upon it, feedback over chosen and forgone acts has to be provided ex-post (Zeelenberg, 1999; Zeelenberg and Pieters, 2007). If ex-post feedback was necessary to trigger regret-averse choices, it would not be so surprising that we did not observe any meaningful pattern of regret aversion. Interestingly, another closely related work by Somasundaram and Diecidue (2017), which adopted the same experimental methodology by BCD, confirmed the predominance of regret aversion in their no-feedback treatment and found that providing immediate choice feedback to subjects could influence their risk and regret attitudes but not as what their theory would predict.[27] However, in Loewenfeld and Zheng (2023)'s test of generalized regret theory, subjects also received immediate feedback on their choices but no significant feedback effect was observed. Put together, the evidence for the effect of feedback on choice seems quite mixed. We conclude that the contradictory findings at the individual level between our study and BCD cannot be explained by these common and different features of the experimental design.

# 8 Conclusion

This paper investigated the empirical properties of RT as a descriptive theory of decision-making under risk. Starting from an experimental design introduced by BCD, the aim was to validate their findings with a larger and more representative sample, and to extend the analysis from financial to health outcomes. Contrary to BCD, we do not find clear evidence in favor of RT. Depending on the precise measure used to classify regret functions as convex or concave, we find that the share of subjects exhibiting a convex regret function can be significantly higher or significantly lower than the share of concave subjects. Moreover, a substantial share of subjects violate the predictions of RT in a specifically designed consistency check. We also find no evidence of systematic differences in regret attitude between financial and health domains. Overall, we conclude that RT does not describe the preferences of the majority of subjects in our sample, and that its predictive accuracy as a theory of decision-making under risk should be questioned.

[27] As the authors also pointed out, the immediate feedback effect they documented could be an outcome of experienced regret instead of anticipated regret.

**Acknowledgments** We thank the TSE ERNA graduates and alumni who participated in the pilot and provided valuable feedback. H.S. thanks Frederik H. Bennhoff for feedback on the manuscript.

**Funding** This study was funded by the French National Research Agency (ANR), grant number ANR-20-CE36-0004-1. Henrik Andersson also acknowledges funding from ANR, France under grant ANR-17-EURE-0010 (Investissements d'Avenir program). The usual disclaimers apply.

**Availability of data and code** The dataset and analysis code can be retrieved from https://drive.google.com/file/d/1Pe-_Q4webvj60e0ZochtlhQlZKRAoA6-/view?usp=share_link.

**Ethics approval** The study was approved by the ethics committee of the Toulouse School of Economics.

**Conflicts of interest** The authors declare no conflicts of interest.

# References

Abdellaoui, M., H. Bleichrodt, and C. Paraschiv (2007). Loss aversion under prospect theory: A parameter-free measurement. *Management Science 53*(10), 1659–1674.

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica 21*(4), 503–546.

Alós-Ferrer, C. and A. Ritschel (2022). Attention and salience in preference reversals. *Experimental Economics 25*(3), 1024–1051.

Arechar, A. A., S. Gächter, and L. Molleman (2018). Conducting interactive experiments online. *Experimental Economics 21*(1), 99–131.

Auer, E. M., T. S. Behrend, A. B. Collmus, R. N. Landers, and A. F. Miles (2021). Pay for performance, satisfaction and retention in longitudinal crowdsourced research. *Plos One 16*(1), e0245460.

Becerra Pérez, M. M., M. Menear, J. C. Brehaut, and F. Legare (2016). Extent and predictors of decision regret about health care decisions: a systematic review. *Medical Decision Making 36*(6), 777–790.

Bell, D. E. (1982). Regret in decision making under uncertainty. *Operations Research 30*(5), 961–981.

Bleichrodt, H., A. Cillo, and E. Diecidue (2010). A quantitative measurement of regret theory. *Management Science 56*(1), 161–175.

Bleichrodt, H. and P. P. Wakker (2015). Regret theory: A bold alternative to the alternatives. *The Economic Journal 125*(583), 493–532.

Bordalo, P., N. Gennaioli, and A. Shleifer (2012). Salience theory of choice under risk. *The Quarterly Journal of Economics 127*(3), 1243–1285.

Brewer, N. T., J. T. DeFrank, and M. B. Gilkey (2016). Anticipated regret and health behavior: A meta-analysis. *Health Psychology 35*(11), 1264.

Carlsson, F. and P. Martinsson (2001). Do hypothetical and actual marginal willingness to pay differ in choice experiments? application to the valuation of the environment. *Journal of Environmental Economics and Management 41*(2), 179–192.

Cubitt, R. P., C. Starmer, and R. Sugden (1998). On the validity of the random lottery incentive system. *Experimental Economics 1*(2), 115–131.

Diecidue, E. and J. Somasundaram (2017). Regret theory: A new foundation. *Journal of Economic Theory 172*, 88–119.

Fréchette, G. R., K. Sarnoff, and L. Yariv (2021). Experimental economics: Past and future. *Unpublished Manuscript.*.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives 19*(4), 25–42.

Gandullia, L. (2019). The price elasticity of warm-glow giving. *Economics Letters 182*, 30–32.

Gilovich, T. and V. H. Medvec (1995). The experience of regret: what, when, and why. *Psychological Review 102*(2), 379.

Guiso, L. (2015). A test of narrow framing and its origin. *Italian Economic Journal 1*(1), 61–100.

Herweg, F. and D. Müller (2021). A comparison of regret theory and salience theory for decisions under risk. *Journal of Economic Theory 193*, 105226.

Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica 47*(2), 263–292.

List, J. A., A. M. Shaikh, and Y. Xu (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics 22*, 773–793.

Loewenfeld, M. and J. Zheng (2023). Uncovering correlation sensitivity in decision making. Technical report, Working Paper.

Loomes, G. (1988). When actions speak louder than prospects. *American Economic Review 78*(3), 463–470.

Loomes, G. and R. Sugden (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal 92*(368), 805–824.

Loomes, G. and R. Sugden (1987a). Some implications of a more general form of regret theory. *Journal of Economic Theory 41*(2), 270–287.

Loomes, G. and R. Sugden (1987b). Testing for regret and disappointment in choice under uncertainty. *The Economic Journal 97*, 118–129.

Paolacci, G., J. Chandler, and P. G. Ipeirotis (2010). Running experiments on amazon mechanical turk. *Judgment and Decision making 5*(5), 411–419.

Richard, T. and V. Baudin (2020). Asymmetric noise and systematic biases: A new look at the trade-off method. *Economics Letters 191*, 109132.

Romano, J. P. and M. Wolf (2005). Stepwise multiple testing as formalized data snooping. *Econometrica 73*(4), 1237–1282.

Savage, L. J. (1951). The theory of statistical decision. *Journal of the American Statistical association 46*(253), 55–67.

Schwartz, B., A. Ward, J. Monterosso, S. Lyubomirsky, K. White, and D. R. Lehman (2002). Maximizing versus satisficing: happiness is a matter of choice. *Journal of Personality and Social Psychology 83*(5), 1178.

Scott, A. J. and G. A. Seber (1983). Difference of proportions from the same survey. *The American Statistician 37*(4a), 319–320.

Smith, R. D. (1996). Is regret theory an alternative basis for estimating the value of healthcare interventions? *Health Policy 37*(2), 105–115.

Somasundaram, J. and E. Diecidue (2017). Regret theory and risk attitudes. *Journal of Risk and Uncertainty 55*(2), 147–175.

Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature 38*(2), 332–382.

Wakker, P. and D. Deneffe (1996). Eliciting von neumann-morgenstern utilities when probabilities are distorted or unknown. *Management Science 42*(8), 1131–1150.

Zeelenberg, M. (1999). Anticipated regret, expected feedback and behavioral decision making. *Journal of Behavioral Decision Making 12*(2), 93–106.

Zeelenberg, M. and R. Pieters (2007). A theory of regret regulation 1.0. *Journal of Consumer Psychology 17*(1), 3–18.

Zheng, J. (2021). Willingness to pay for reductions in health risks under anticipated regret. *Journal of Health Economics 78*, 102476.

# Appendix A: Survey instructions, scenario descriptions and lottery choices

Figure A1. Welcome screen

## Behavioral determinants of individual risk preferences

This study is part of a project on the behavioral determinants of individual risk preferences at the Toulouse School of Economics, France. Your answers will be used for research purposes and provide insights on how individuals make choices in situations when the outcome is uncertain, which can be useful for policy makers when designing policies involving risks. We would like to emphasize that your contribution is very important. We are interested in answers both from individuals who are certain about their answers as those who are less certain.

Participation in the study is voluntary, and you can withdraw at any time. In total, the survey should take around 15-25 minutes to complete. All responses will be handled completely anonymously.

If you have any questions regarding the survey and how the results can be used you can contact us at henrik.andersson@tse-fr.eu.

If you consent to participate in the survey click on the "Next" button below. If you do not consent you leave the survey by closing this window.

Henrik Andersson and Henrik Scholtz
Toulouse School of Economics

This survey is anonymous.

The record of your survey responses does not contain any identifying information about you, unless a specific survey question explicitly asked for it.

If you used an identifying token to access this survey, please rest assured that this token will not be stored together with your responses. It is managed in a separate database and will only be updated to indicate whether you did (or did not) complete this survey. There is no way of matching identification tokens with survey responses.

Next

Figure A2.1. Scenario description in scenario H (health outcomes)

Imagine you have been diagnosed with a rare medical condition that, if left untreated, will lead to your immediate death. Your doctor proposes two different types of treatment - treatments A and B - both of which will increase your remaining life expectancy as described in the questions.

The effect of both treatments is uncertain. With some known probability your body is more or less receptive to each treatment. Please assume that your remaining life expectancy in each case can accurately be predicted. Further assume that the treatment is costless for you, and that the only effect of the medical condition and of the treatment is to change your remaining life expectancy. That is, neither the condition nor the treatment has any effect on your health while you are still alive.

In each of the following questions, choose whichever option you would prefer for yourself. There are no right or wrong answers.

Next

Figure A2.2. Scenario description in scenario U (unincentivized financial outcomes)

In what is to follow, you will face pairwise choices between different lotteries. The outcome of each pair of lotteries is randomly determined based on a random number generator that draws any number from 1 to 100 with equal probability. Depending on this randomly drawn number, the lotteries yield different payoffs as described in the questions.

In each of the following questions, please choose whichever lottery you would prefer. Please note that the choices made are hypothetical and that you won't receive any additional money based on your choices. There are no right or wrong answers.

Next

Figure A2.3. Scenario description in scenario I (incentivized financial outcomes)

In what is to follow, you will face pairwise choices between different lotteries. The outcome of each pair of lotteries is randomly determined based on a random number generator that draws any number from 1 to 100 with equal probability. Depending on this randomly drawn number, the lotteries yield different payoffs as described in the questions.

In each of the following questions, please choose whichever lottery you would prefer. There are no right or wrong answers. Please note that for 10% of all participants, one of the lotteries will randomly be chosen to actually be paid out. You will be informed at the end of this survey whether you are among the 10% of participants who receive additional payments based on their lottery choices. The payment will be added to your compensation for completing the survey on Amazon MTurk. Even though you will make many decisions, only one of these will end up affecting your earnings, but you will not know in advance which decision your payment will be based on, if you are chosen for additional payment.

***Notice: The payment can only be made if you successfully register your completion of the survey on Amazon MTurk.***

Next

Figure A3.1. Practice question and feedback in scenarios U and I

*This is a practice question meant to get you accustomed to the types of choices you will have to make throughout the survey. You will receive feedback on the outcome of your choice once you have selected a lottery.

The table below summarizes your payoffs for each of the two lotteries, depending on the random number that will be drawn. For example, if you choose lottery A and the number 73 is drawn, your payoff is $6. If you had chosen lottery B instead, it would have been $5. If you choose lottery B and the number 32 is drawn, your payoff is $11. If you had chosen lottery A instead, it would have been $10.

| | **Numbers 1-50 (p = 50%)** | **Numbers 51-100 (p = 50%)** |
|---|---|---|
| **Lottery A** | $10 | $6 |
| **Lottery B** | $11 | $5 |

Which of the two lotteries would you prefer?

Choose one of the following answers

- ○ Lottery A
- ○ Lottery B

Next

*This is a practice question meant to get you accustomed to the types of choices you will have to make throughout the survey. You will receive feedback on the outcome of your choice once you have selected a lottery.

The table below summarizes your payoffs for each of the two lotteries, depending on the random number that will be drawn. For example, if you choose lottery A and the number 73 is drawn, your payoff is $6. If you had chosen lottery B instead, it would have been $5. If you choose lottery B and the number 32 is drawn, your payoff is $11. If you had chosen lottery A instead, it would have been $10.

| | **Numbers 1-50 (p = 50%)** | **Numbers 51-100 (p = 50%)** |
|---|---|---|
| **Lottery A** | $10 | $6 |
| **Lottery B** | $11 | $5 |

Which of the two lotteries would you prefer?

Choose one of the following answers

- ● Lottery A
- ○ Lottery B

You have chosen lottery A. Suppose the randomly drawn number is 65. Then your payoff is $6. Had you chosen lottery B, it would have been $5 instead.

Next

Figure A3.2. Practice question and feedback in scenario H

*This is a practice question meant to get you accustomed to the types of choices you will have to make throughout the survey. You will receive feedback on the outcome of your choice once you have selected a treatment.

The table below summarizes your life expectancy for each of the two treatments depending on your body's receptiveness to the treatments. Your body can be more or less receptive, which each happens with 50% probability. Notice that if you are receptive to treatment A, you are also receptive to treatment B and vice versa. For example, if you choose treatment A and your body is less receptive, your life expectancy is 6 months. If you had chosen treatment B instead, it would have been 5 months. If you choose treatment B and your body is more receptive, your life expectancy is 11 months. If you had chosen treatment A instead, it would have been 10 months.

| | **More receptive to treatment (p = 50%)** | **Less receptive to treatment (p = 50%)** |
|---|---|---|
| **Treatment A** | 10 months | 6 months |
| **Treatment B** | 11 months | 5 months |

Which of the two treatments would you prefer?

Choose one of the following answers

○ Treatment A

○ Treatment B

Next

*This is a practice question meant to get you accustomed to the types of choices you will have to make throughout the survey. You will receive feedback on the outcome of your choice once you have selected a treatment.

The table below summarizes your life expectancy for each of the two treatments depending on your body's receptiveness to the treatments. Your body can be more or less receptive, which each happens with 50% probability. Notice that if you are receptive to treatment A, you are also receptive to treatment B and vice versa. For example, if you choose treatment A and your body is less receptive, your life expectancy is 6 months. If you had chosen treatment B instead, it would have been 5 months. If you choose treatment B and your body is more receptive, your life expectancy is 11 months. If you had chosen treatment A instead, it would have been 10 months.

| | **More receptive to treatment (p = 50%)** | **Less receptive to treatment (p = 50%)** |
|---|---|---|
| **Treatment A** | 10 months | 6 months |
| **Treatment B** | 11 months | 5 months |

Which of the two treatments would you prefer?

Choose one of the following answers

● Treatment A

○ Treatment B

You have chosen treatment A. Suppose it turns out that you are more receptive to the treatment. Then your life expectancy is 10 months. Had you chosen treatment B, it would have been 11 months instead.

Next

Figure A4.1. Comprehension questions for scenarios U and I. This figure shows the initial questions. For scenario H, the wording and payoffs were adjusted accordingly (life expectancy instead of monetary payoffs).

The following questions are meant to ensure your understanding of the lottery choice task. You will receive feedback on your answers on the next page.

As described in the practice question, the table below summarizes your payoffs for each of the two lotteries, depending on the random number that will be drawn. For example, if you choose lottery A and the number 73 is drawn, your payoff is $8. If you had chosen lottery B instead, it would have been $6. If you choose lottery B and the number 32 is drawn, your payoff is $14. If you had chosen lottery A instead, it would have been $12.

| | **Numbers 1-60** (p = 60%) | **Numbers 61-100** (p = 40%) |
|---|---|---|
| **Lottery A** | $12 | $8 |
| **Lottery B** | $14 | $6 |

*Suppose you chose lottery A. What will your payoff be?

Choose one of the following answers

- $14 with 60% probability, $6 with 40% probability.
- $12 with 60% probability, $8 with 40% probability.
- $14 with 60% probability, $12 with 60% probability.

*Suppose the randomly drawn number is 29. Which of the following is correct?

Choose one of the following answers

- Lottery A yields a payoff of $14 and lottery B yields a payoff of $12.
- Lottery A yields a payoff of $8 and lottery B yields a payoff of $6.
- Lottery A yields a payoff of $12 and lottery B yields a payoff of $14.

*Which of the following is correct?

Choose one of the following answers

- Lottery A yields a higher payoff than lottery B with 40% probability.
- Lottery A always yields the same payoff.
- Lottery B always yields a higher payoff than lottery A.

Next

Figure A4.2. Feedback when the answer to one of the comprehension questions was wrong (scenarios U and I; scenario H adjusted accordingly)

You have answered at least one of the questions incorrectly. You will find feedback on your answers below. Please read the questions again carefully and try to answer them correctly.

As described in the practice question, the table below summarizes your payoffs for each of the two lotteries, depending on the random number that will be drawn. For example, if you choose lottery A and the number 73 is drawn, your payoff is $8. If you had chosen lottery B instead, it would have been $6. If you choose lottery B and the number 32 is drawn, your payoff is $14. If you had chosen lottery A instead, it would have been $12.

| | **Numbers 1-60 (p = 60%)** | **Numbers 61-100 (p = 40%)** |
|---|---|---|
| **Lottery A** | $12 | $8 |
| **Lottery B** | $14 | $6 |

Suppose you chose lottery A. What will your payoff be?

You answered this question correctly. If you choose lottery A, your payoff is $12 with 60% probability and $8 with 40% probability.

Suppose the randomly drawn number is 29. Which of the following is correct?

You answered this question correctly. If the randomly drawn number is 29, lottery A yields a payoff of $12 and lottery B yields a payoff of $14.

*Which of the following is correct?

You answered this question incorrectly. Please read the question again carefully and try to answer it correctly.

Choose one of the following answers

- Lottery A yields a higher payoff than lottery B with 40% probability.
- Lottery A always yields the same payoff.
- Lottery B always yields a higher payoff than lottery A.

Next

Figure A4.3. Feedback when all comprehension questions were answered correctly (scenarios U and I; scenario H adjusted accordingly)

You have answered all questions correctly. You will find feedback on your answers below.

As described in the practice question, the table below summarizes your payoffs for each of the two lotteries, depending on the random number that will be drawn. For example, if you choose lottery A and the number 73 is drawn, your payoff is $8. If you had chosen lottery B instead, it would have been $6. If you choose lottery B and the number 32 is drawn, your payoff is $14. If you had chosen lottery A instead, it would have been $12.

| | **Numbers 1-60 (p = 60%)** | **Numbers 61-100 (p = 40%)** |
|---|---|---|
| **Lottery A** | $12 | $8 |
| **Lottery B** | $14 | $6 |

Suppose you chose lottery A. What will your payoff be?

You answered this question correctly. If you choose lottery A, your payoff is $12 with 60% probability and $8 with 40% probability.

Suppose the randomly drawn number is 29. Which of the following is correct?

You answered this question correctly. If the randomly drawn number is 29, lottery A yields a payoff of $12 and lottery B yields a payoff of $14.

Which of the following is correct?

You answered this question correctly. Lottery A yields a higher payoff than lottery B with 40% probability.

***Now that you are acquainted with the structure of the lottery choice questions, you can progress to the main part of the survey. Please pay close attention to the payoffs and probabilities, since many of the values may stay the same from one question to the next while some others change.***

Next

Figure A5.1. Exemplary lottery choice question in the utility elicitation

*As before, the table below summarizes your payoffs for each of the two lotteries, depending on the random number that will be drawn.

| | Numbers 1-50 (p = 50%) | Numbers 51-100 (p = 50%) |
|---|---|---|
| **Lottery A** | $16 | $12 |
| **Lottery B** | $24 | $10 |

Which of the two lotteries would you prefer?

Choose one of the following answers

- Lottery A
- Lottery B

Next

Figure A5.2 Exemplary lottery choice question in the regret elicitation. Changes in probabilities were highlighted both in the text and the payoff matrix

*The table below summarizes your payoffs for each of the two lotteries, depending on the random number that will be drawn. Notice that **now**, the probability of receiving the payoff in the left column of each lottery is **25%**, while the probability of receiving the payoff in the right column is **75%**.

| | Numbers 1-25 (p = 25%) | Numbers 26-100 (p = 75%) |
|---|---|---|
| **Lottery A** | $26 | $16 |
| **Lottery B** | $23.5 | $20 |

Which of the two lotteries would you prefer?

Choose one of the following answers

- Lottery A
- Lottery B

Next

# Appendix B: Psychological regret scales

Figure B1. General regret scale from Schwartz et al. (2002)

*For each of the following statements, please indicate to what extent you agree or disagree (1 = completely disagree, 7 = completely agree).

| | 1 (completely disagree) | 2 | 3 | 4 | 5 | 6 | 7 (completely agree) |
|---|---|---|---|---|---|---|---|
| Whenever I make a choice, I'm curious about what would have happened if I had chosen differently. | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Whenever I make a choice, I try to get information about how the other alternatives turned out. | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| If I make a choice and it turns out well, I still feel like something of a failure if I find out that another choice would have turned out better. | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| When I think about how I'm doing in life, I often assess opportunities I have passed up. | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Once I make a decision, I don't look back. | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

Next

Figure B2. Financial regret scale adapted from Guiso (2015)

How would you react if you found yourself in the following situation? Two years ago a friend who is knowledgeable about finance recommended that you undertake an investment which, on the basis of the information available to him then, had good chances of success.

*You chose not to make the investment. Meanwhile, the value of the investment has more than doubled and had you made it you could have made a big gain. In such circumstances, today how would you feel about your decision? (1 means 'feel no regret' and 7 means 'deeply regret not having made the investment') :

| | 1 (no regret) | 2 | 3 | 4 | 5 | 6 | 7 (strong regret) |
|---|---|---|---|---|---|---|---|
| | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

*Now think of another situation. You invested a significant amount in the investment that was recommended. Meanwhile market conditions have deteriorated and your investment has lost half of its value. In such circumstances, today how would you feel about your decision? (1 means 'feel no regret' and 7 means 'deeply regret having made the investment') :

| | 1 (no regret) | 2 | 3 | 4 | 5 | 6 | 7 (strong regret) |
|---|---|---|---|---|---|---|---|
| | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

Next

Figure B3. Health regret scale in the style of Guiso (2015)

How would you react if you found yourself in the following situation? Two years ago your doctor diagnosed you with a rare condition and recommended a new treatment which, on the basis of the information available to him then, had good chances of success.

*You chose to decline the treatment. Meanwhile, clinical tests have shown that it is very effective and had you accepted it you could have improved your health substantially. In such circumstances, today how would you feel about your decision? (1 means 'feel no regret' and 7 means 'deeply regret not having taken the treatment') :

| | 1 (no regret) | 2 | 3 | 4 | 5 | 6 | 7 (strong regret) |
|---|---|---|---|---|---|---|---|
| | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

*Now think of another situation. You accepted the treatment that was recommended. However, your health condition has not improved and you suffered from serious side effects. In such circumstances, today how would you feel about your decision? (1 means 'feel no regret' and 7 means 'deeply regret having taken the treatment') :

| | 1 (no regret) | 2 | 3 | 4 | 5 | 6 | 7 (strong regret) |
|---|---|---|---|---|---|---|---|
| | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

Next

# Appendix C: Median and mean utility functions

Figure C1.1. Median utility function in scenario H

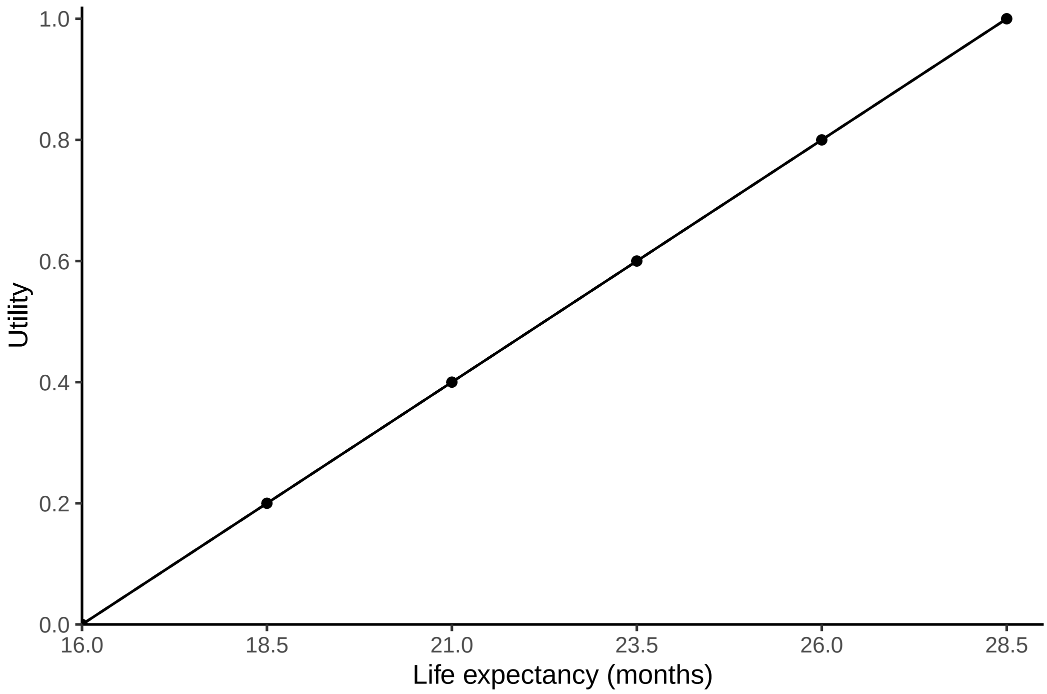



Figure C1.2. Median utility function in scenario U

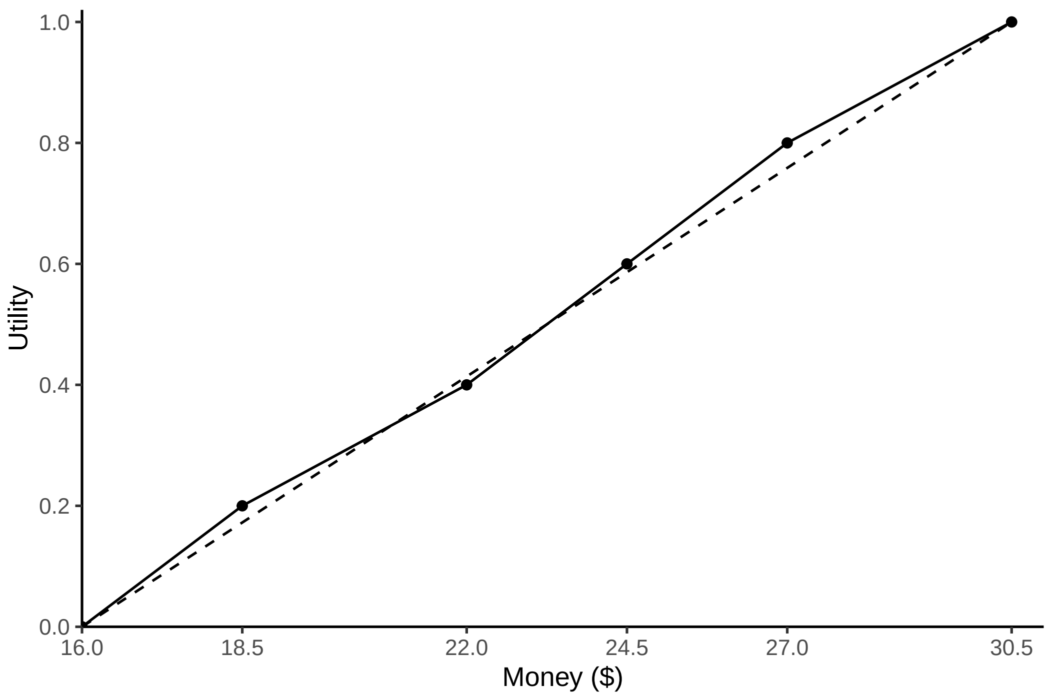


Figure C1.3. Median utility function in scenario I

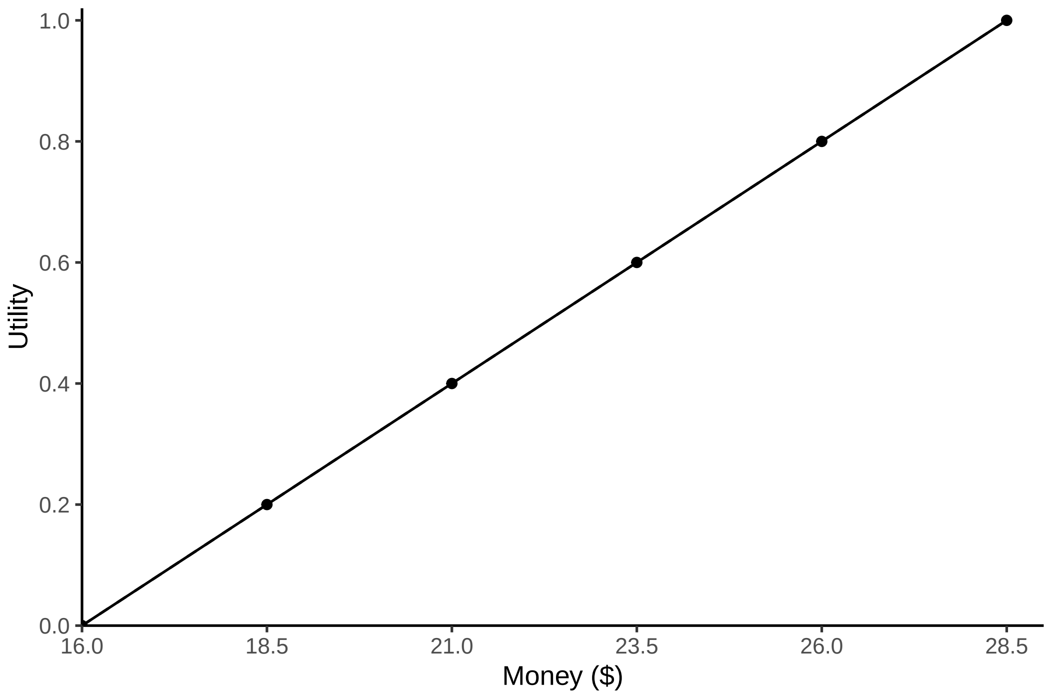



Figure C2.1. Mean utility function in scenario H

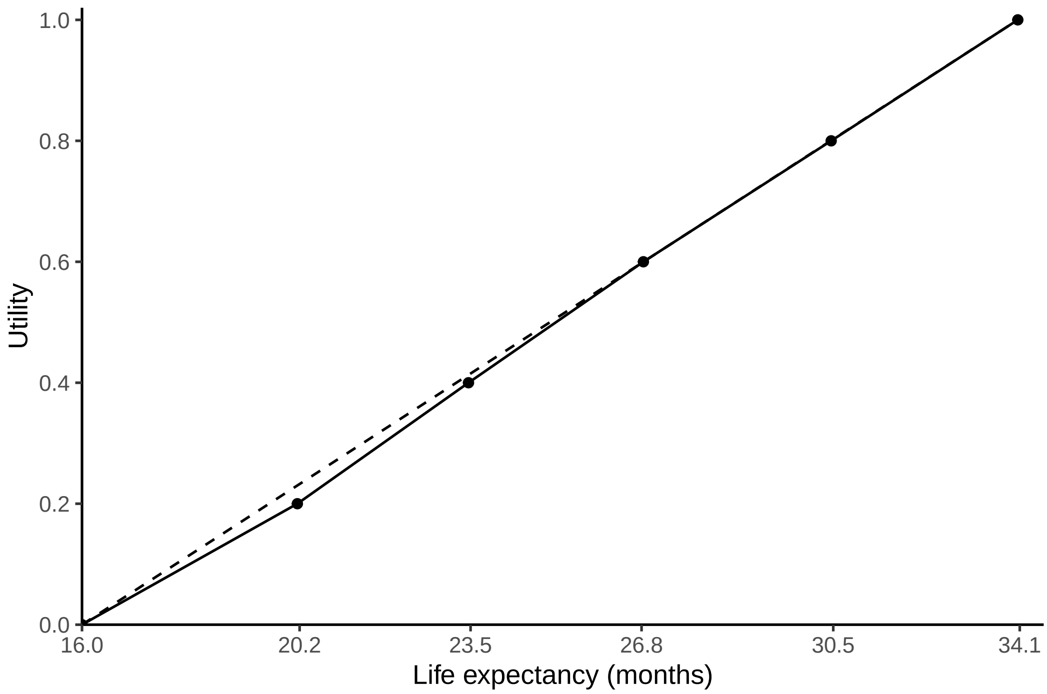


Figure C2.2. Mean utility function in scenario U

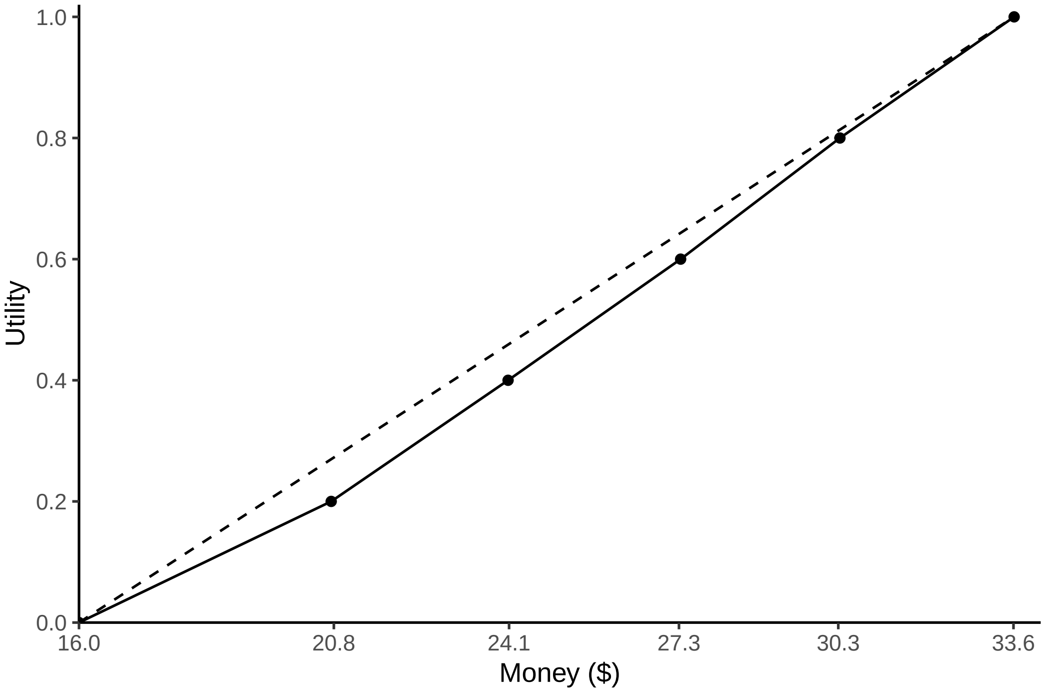
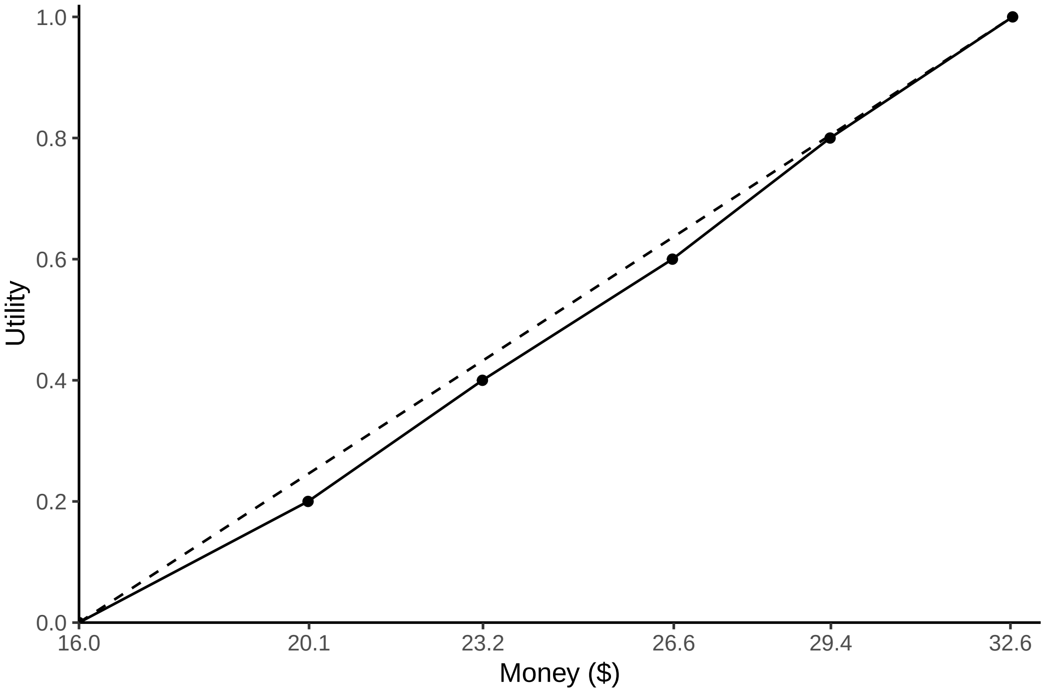



Figure C2.3. Mean utility function in scenario I



# Appendix D: Median and mean regret functions

Figure D1.1. Median regret function in scenario H

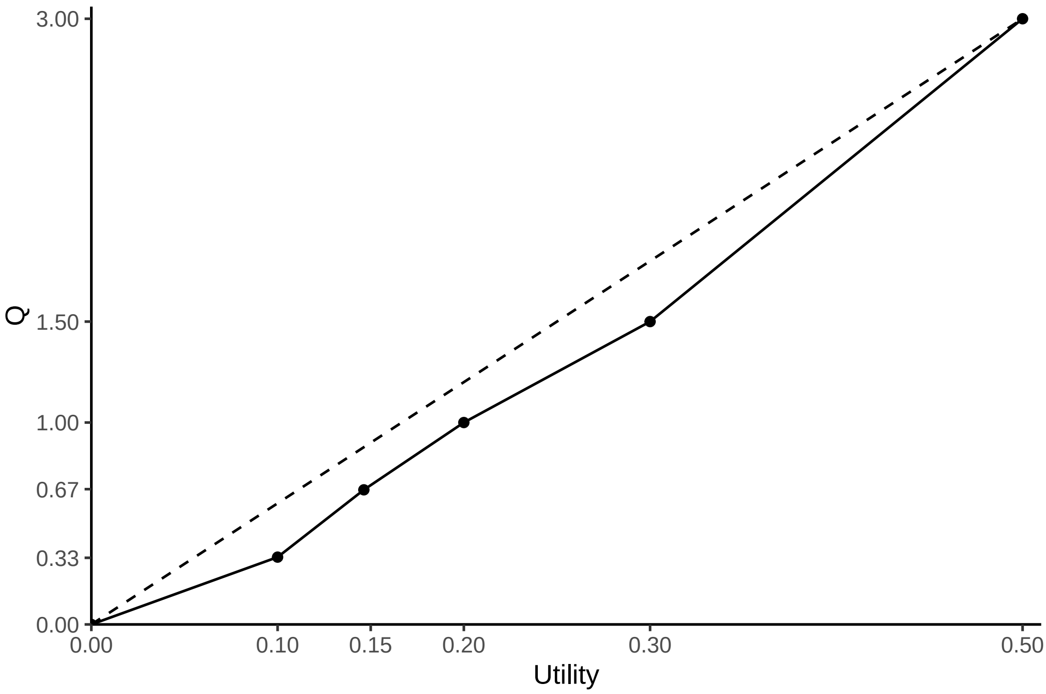



Figure D1.2. Median regret function in scenario U

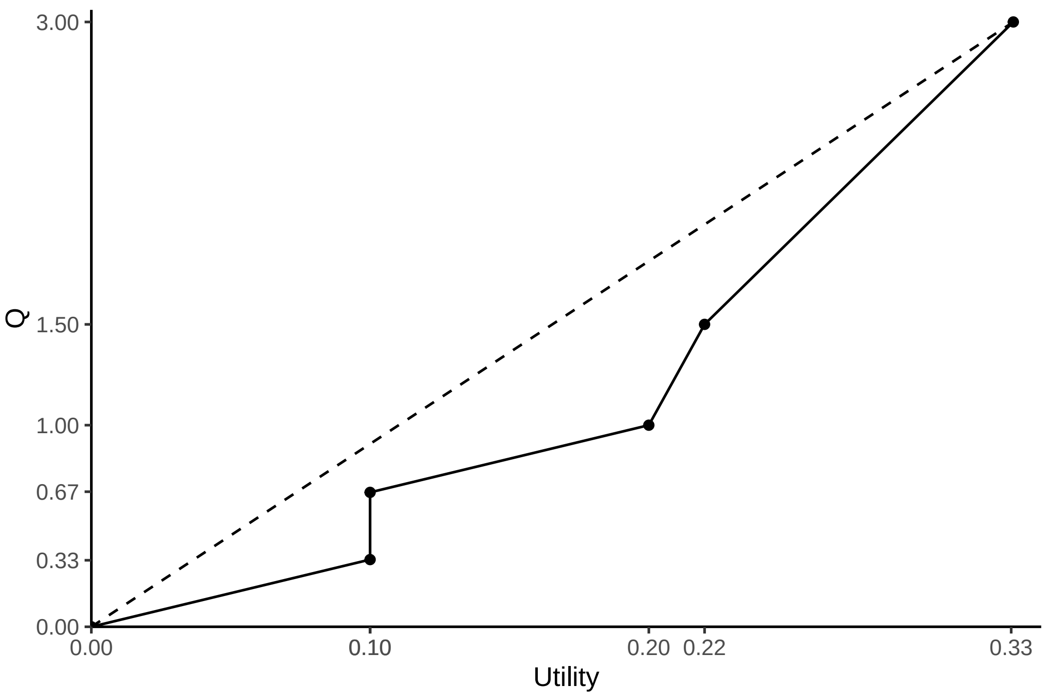


Figure D1.3. Median regret function in scenario I

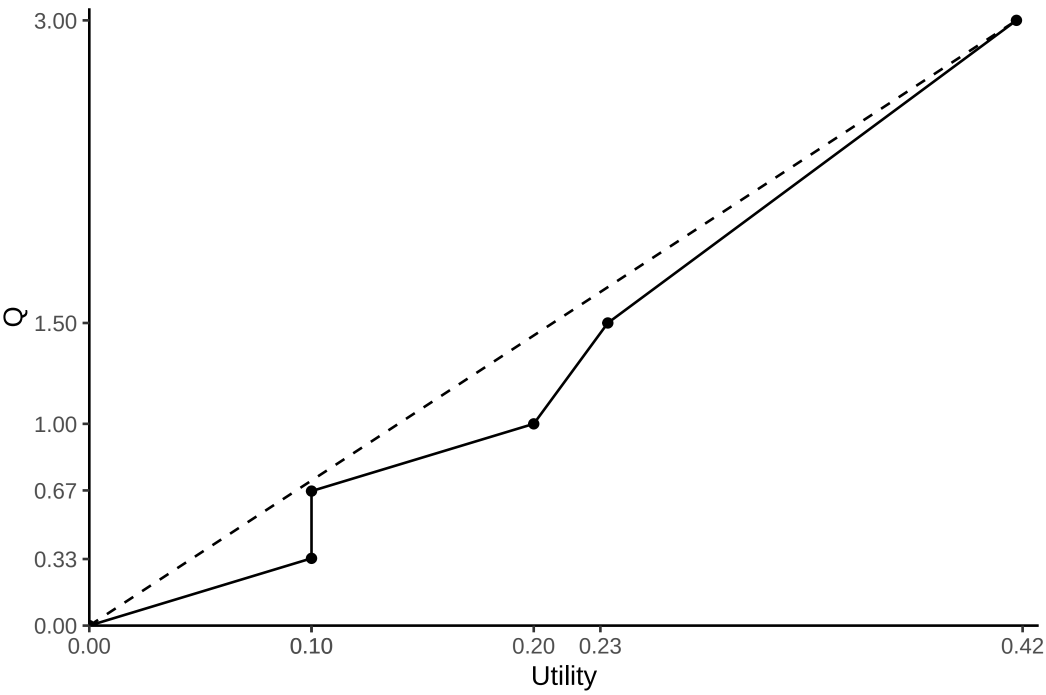



Figure D2.1. Mean regret function in scenario H

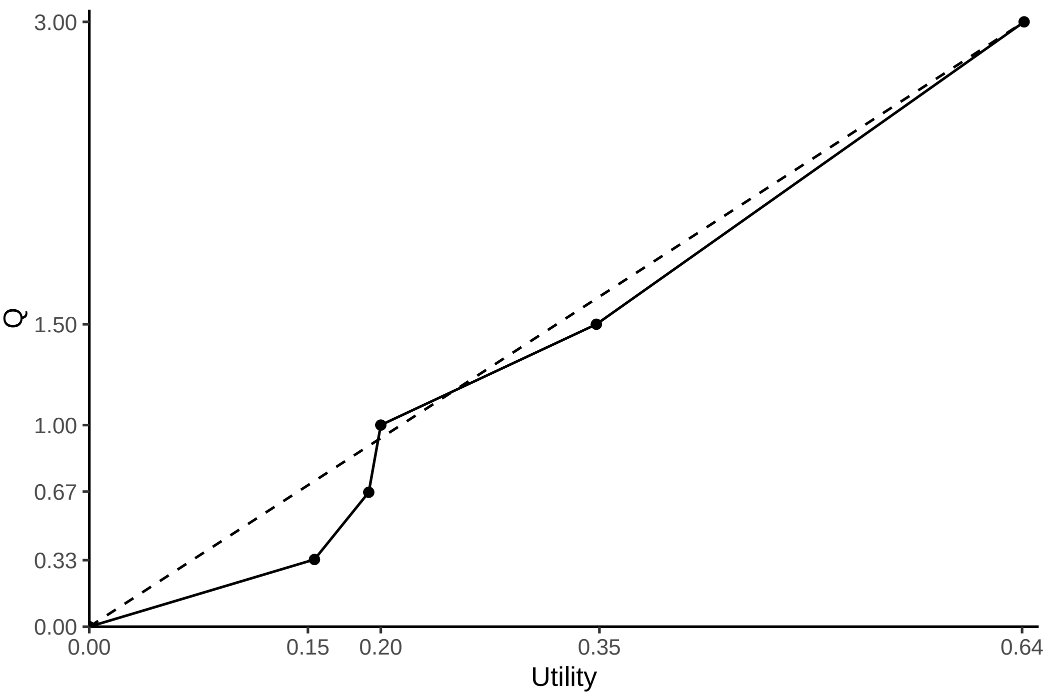


Figure D2.2. Mean regret function in scenario U

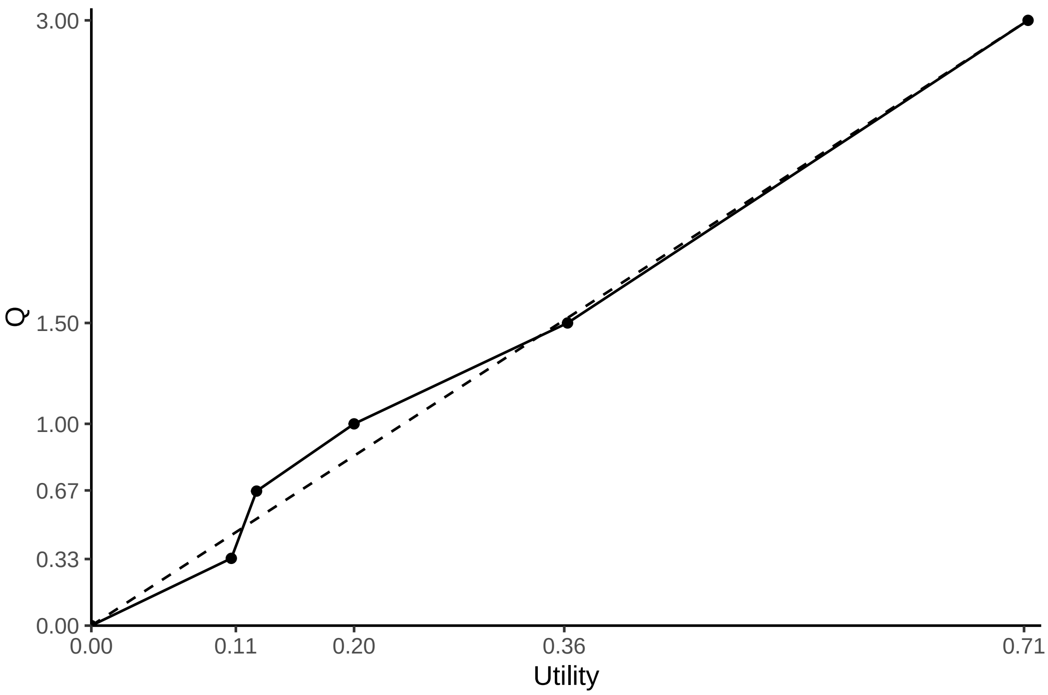
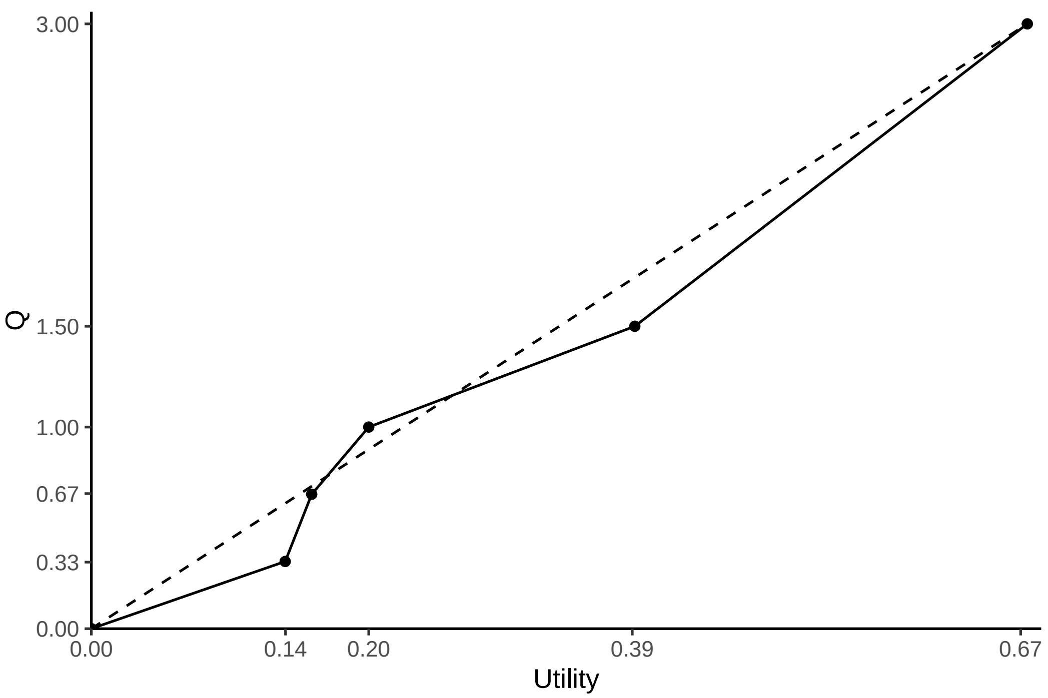



Figure D2.3. Mean regret function in scenario I



# Appendix E: Regret sensitivity for different inclusion criteria

## E.1 Unrestricted sample

We first present results similar to those in section 6.3 for an "unrestricted" sample. That is, we include all subjects with well-defined values throughout the utility and regret elicitation and who correctly answered the comprehension questions. This leaves exactly the 455 subjects that were analyzed in the utility part.

All results remain qualitatively very similar to those presented in section 6.3 (see Table E1). Based on the area under the curve, we would conclude that significantly more subjects exhibit convex as opposed to concave regret functions, with an aggregate ratio of about 4:3. The slope evolution criterion paints a different picture, with a ratio of about 5:2 in favor of concave regret functions using our default cutoff value of 14 and strongly significant differences in the same direction for any cutoff value. The overlap of both measures of

Table E1. Classification of individual regret functions using the "unrestricted" sample

(a) based on the area under the curve

| | Convex | | Concave | Total |
|---|---|---|---|---|
| H | 84 | $\overset{**}{\longleftrightarrow}$ | 53 | 137 |
| U | 81 | | 79 | 160 |
| I | 95 | $\overset{**}{\longleftrightarrow}$ | 63 | 158 |
| Total | 260 | $\overset{***}{\longleftrightarrow}$ | 195 | 455 |

(b) based on the slope evolution criterion (cutoff = 14)

| | Convex | | Concave | Unclassified | Total |
|---|---|---|---|---|---|
| H | 23 | $\overset{**}{\longleftrightarrow}$ | 42 | 72 | 137 |
| U | 16 | $\overset{***}{\longleftrightarrow}$ | 51 | 93 | 160 |
| I | 19 | $\overset{***}{\longleftrightarrow}$ | 51 | 88 | 158 |
| Total | 58 | $\overset{***}{\longleftrightarrow}$ | 144 | 253 | 455 |

(c) based on the slope evolution criterion (different cutoffs, pooled across scenarios)

| Cutoff value | Convex | | Concave | $p \leq 0.05$ in scenario |
|---|---|---|---|---|
| 11 | 117 | $\overset{***}{\longleftrightarrow}$ | 292 | H, U, I |
| 12 | 100 | $\overset{***}{\longleftrightarrow}$ | 246 | H, U, I |
| 13 | 74 | $\overset{***}{\longleftrightarrow}$ | 194 | H, U, I |
| 14 | 58 | $\overset{***}{\longleftrightarrow}$ | 144 | H, U, I |
| 15 | 38 | $\overset{***}{\longleftrightarrow}$ | 94 | H, U, I |
| 16 | 24 | $\overset{***}{\longleftrightarrow}$ | 50 | U, I |

Table E2. Overlap of both measures of regret curvature in the "unrestricted" sample

(a) slope cutoff = 14

| | Convex slope | Unclassified slope | Concave slope |
|---|---|---|---|
| Convex area | 58 | 160 | 42 |
| Concave area | 0 | 93 | 102 |

(b) slope cutoff = 11

| | Convex slope | Unclassified slope | Concave slope |
|---|---|---|---|
| Convex area | 110 | 30 | 120 |
| Concave area | 7 | 16 | 172 |

regret curvatures remains moderately strong, as shown in Table E2. We therefore confirm the analysis from the main text in the larger, unrestricted sample.

## E.2 Restricted sample

We now present analogous results for a strongly restricted sample. Namely, we only include subjects exhibiting increasing regret functions without any margin of error. In addition, we require that their choice in the consistency question is as predicted by regret theory. This leaves us with only 74 subjects.

The very small remaining sample size evidently leaves us with low power to detect any difference between the shares of convex and concave subjects (see Table E3 and E4). Yet the patterns remain qualitatively similar to those we observe with larger, less restricted samples. A modest majority of subjects are labeled convex based on the area under curve, while the opposite is true when using the slope evolution criterion. None of the differences are significant, neither at the aggregate level nor within any scenario. Even when restricting attention to the subjects that ex ante seem most likely to behave according to regret theory, we do not find evidence of systematically convex regret functions. This reinforces the argument made in the main text.

Table E3. Classification of individual regret functions using the strongly restricted sample

(a) based on the area under the curve

| | Convex | Concave | Total |
|---|---|---|---|
| H | 13 | 11 | 24 |
| U | 16 | 9 | 25 |
| I | 13 | 12 | 25 |
| Total | 42 | 32 | 74 |

(b) based on the slope evolution criterion (cutoff = 14)

| | Convex | Concave | Unclassified | Total |
|---|---|---|---|---|
| H | 7 | 8 | 9 | 24 |
| U | 6 | 8 | 11 | 25 |
| I | 5 | 10 | 10 | 25 |
| Total | 18 | 26 | 30 | 74 |

(c) based on the slope evolution criterion (different cutoffs, pooled across scenarios)

| Cutoff value | Convex | Concave | $p \leq 0.05$ in scenario |
|---|---|---|---|
| 11 | 28 | 39 | - |
| 12 | 26 | 33 | - |
| 13 | 20 | 29 | - |
| 14 | 18 | 26 | - |
| 15 | 15 | 19 | - |
| 16 | 8 | 13 | - |

Table E4. Overlap of both measures of regret curvature in the strongly restricted sample

(a) slope cutoff = 14

| | Convex slope | Unclassified slope | Concave slope |
|---|---|---|---|
| Convex area | 18 | 22 | 2 |
| Concave area | 0 | 8 | 24 |

(b) slope cutoff = 11

| | Convex slope | Unclassified slope | Concave slope |
|---|---|---|---|
| Convex area | 27 | 4 | 11 |
| Concave area | 1 | 3 | 28 |