Online Supplements File for Ph.D. dissertation

Title: Decisions about health - Behavioral experiments in health with applications to understand and improve health state valuation

Author: S.A. (Stefan) Lipman

Supervisors: A.E. (Arthur) Attema & W.B.F. (Werner) Brouwer

File description: Various supplemental methodological details, analyses and results relevant to the Chapters of this dissertation. These Online Supplements will not be printed with the main file. This file also comes with a separate reference list (page X). All Tables and Figures are labelled as follows: Table/Figure OX.Y.Z, where O signifies this is an online Table, X is the Chapter the table or figure is associated with and Y is the Appendix in which it is located and Z is is the consecutive Figure or Table number. Figures or Tables without an O starting them are, thus, figures and tables from the main dissertation.

Appendix Chapter 2

The Online Supplements for this Chapter contains a total of four appendices that provide additional information on the gambles used in Chapter 2 (Appendix A), additional theoretical background describing the role of initial health (Appendix B), discuss the definition of Rabin's Paradox when subjects are indifferent for moderate stake gambles (Appendix C) and present additional analyses and results with regard to the effect of demographic variables and stability of RP preferences across individual and societal outcomes (Appendix D).

Abbreviations:

RP	Rabin's Paradox
EU	Expected Utility
MSG	Moderate stakes gamble
REA	Rabin's empirical assumption
RP1 – RP6	Questions used for demonstrating RP

Appendix A: Instructions used for Rabin Paradox Gamble-Pairs

General Instructions

In this part of the questionnaire we will offer you different gambles, in which you can gain or lose health outcomes. We ask you whether you would play this gamble. You can answer either positively ('Yes, I would') or negatively ('No, I would not'). All gambles involve health outcomes, and we ask you to take different perspectives throughout this short survey. Furthermore, each gamble has the same chance of generating success or failure, namely 50%. The probability of success is not affected by any personal characteristics, and can be viewed as the result of flipping a single coin.

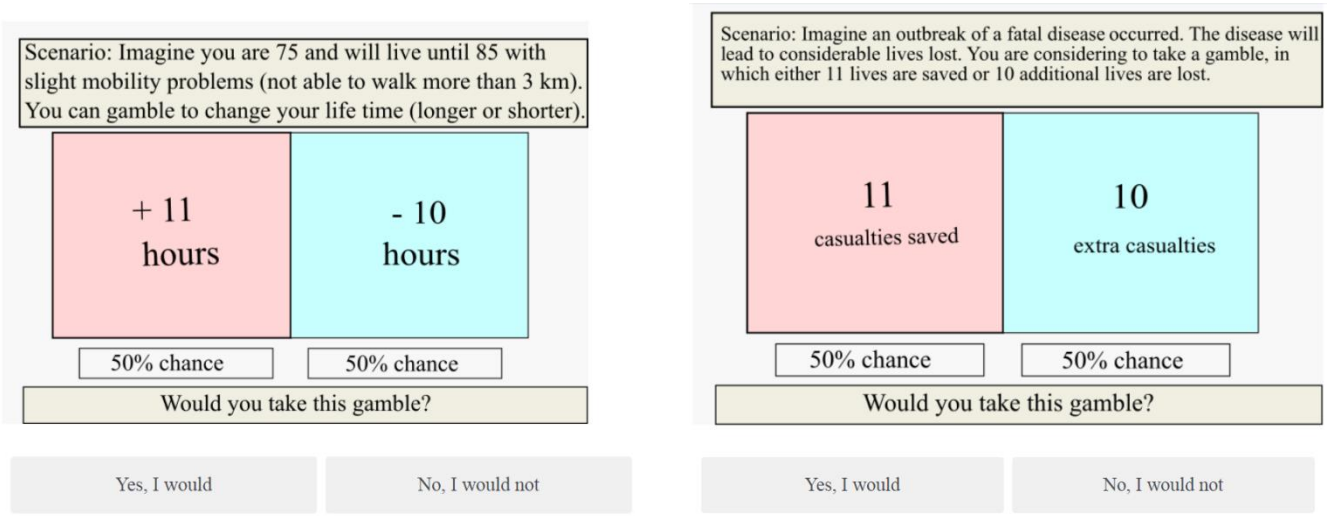
Individual perspective (gamble-pairs: RP1, RP2 and RP3)

In this part of the questionnaire, we ask you to imagine that the outcomes of the gamble will affect you personally. In other words, the health gains or losses that result from the gamble will affect your life time and/or quality of life. For each gamble, it will be clarified when the outcomes occur. Your task is to indicate if you would play this gamble, given that the outcomes affect your personal situation.

Societal perspective (gamble-pairs: RP4, RP5, and RP6)

For this part of the questionnaire, we ask you to take a societal perspective. In other words, you are to decide like a policy maker in health care, who cares for the outcomes of everyone in society together. Imagine that you do not know exactly how the outcomes will be distributed, nor do you know your own position in society.

Figure O.2.A.1: Example item for individual perspective (left panel: RP3 – MSG) and societal perspective (right panel: RP6 – MSG)



Appendix B. Formalizing the role of initial wealth and health

Wakker (2010) provided a formalization of the role of initial wealth in Rabin's (2000) critique on EU based on calibration of moderate stake gambles. This formalization is instrumental to understand how plausible risk aversion over small gambles leads to implausible risk aversion over calibrated gambles – which could be accommodated by reference-dependence. Moreover, by extending this framework to health outcomes, we are able to allude to the plausibility of constant risk aversion over 'common' levels of $\mathcal{H}$. Importantly, health outcomes, as opposed to monetary amounts, have several qualities that may hamper the generality of RP. We try to allude to these briefly in this Appendix. Throughout, we will use the notation defined in Chapter 2, and add to it where necessary.

EU is reference-independent, which for health outcomes indicates that outcomes are evaluated over final health $F_h \in \mathcal{H}$. When considering gambles, such as the moderate stake gamble $\mathcal{g}_p\ell$, the possible corresponding levels of F_h are: $I_h + \mathcal{g}$ or $I_h - \ell$, signifying that outcomes designate changes from I_h. A crucial assumption of EU is that preferences are not affected by adding some amount to $I_h$and subtracting it from all outcomes in a gamble, as this does not affect final health. That is, EU over health outcomes ($\mathcal{H}$) holds if preferences over these outcomes can be represented by $U(\mathcal{H})$. Preferences maximize the expectation of $U(\mathcal{H})$ over final health, where I_h is typically not denoted, as it is assumed constant to the agent and is usually unknown. In economics, this rescaling has been standard for decades (Wakker, 2010).

The crucial assumption to obtain RP is risk aversion for common levels of I_h, which Wakker (2010) refers to as Rabin's empirical assumption (REA). That is, if $0 \succ \mathcal{g}_{0.5}\ell$ at I_h, then for some 'common' range of possible levels of $I_h \in \mathcal{H}$, this preference should remain constant. As in Wakker (2010), we are not able to completely formalize what 'common levels for I_h' refer to, but any range of 12,000 of such 'common levels' for the health outcomes under consideration (see Table O.2.B.1) suffices for our definition of RP. In our experiment, we observed such preferences for $\mathcal{g} = 11, p = \frac{1}{2},$ and $\ell = -11$. Table O.2.B.1 suggests that such possible ranges appear to be conceivable and realistic for health outcomes and may occur in practice (e.g. for individual outcomes we have a difference of 1.4 year in life expectancy). By extending this assumption of constant risk aversion to health, i.e. for many 'common levels of I_h' we thus assume that: a) individuals with different 'common levels of I_h' will turn down gambles[1] of the form $\mathcal{g}_{0.5}\ell$, at least on a range of up to 12,000 $\in \mathcal{H}$, b) as in Rabin (2000) we assume that if this between-subject risk aversion is likely for 'common levels of I_h', it shall also hold for a single subject at different levels of I_h, again at least on a range of 12,000 $\in \mathcal{H}$.

Table O.2.B.1: Ranges of 12,000 for each unit of $\mathcal{H}$ under consideration in RP gamble-pairs with real-life examples of comparable magnitude

Gamble-pair	Ranges of 12,000 under consideration
Individual	

[1] Our paper provides some evidence for this empirical claim, as even in our homogeneous sample subjects turning down the gamble will have varying expectations about their length of life (see for example Pentek et al., 2014), and will thus have different I_h.

RP1-3	Changes in life duration of 1.4 life year (12,000 hours)
Societal	
RP4	12,000 disease cases
RP5	12,000 casualties for an unspecified disease
RP6	12,000 healthy life years

Under these assumptions, as Wakker (2010) concludes in his Chapter 8.6, if EU in terms of final health holds, this implies that for many levels I_h and many non-zero outcomes $x \in \mathcal{H}$ we should have:

$$x \succ (0.5: x + 11, 0.5, x - 10).$$

If we, as Rabin (2000) assumed, take $U(\mathcal{H})$ to be strictly increasing and concave, we have:

$U(x + 11) - U(x) < U(x) - U(x - 10)$, signifying that the marginal utility on the interval [x, x+11] is smaller than on the domain [x-10, x]. This implies that on the domain [x-10, x+11] of $\mathcal{H}$ of length 21, U' drops by a factor of at least 10/11 (Wakker, 2010). Extending this to many levels I_h and many non-zero outcomes x, we can calibrate this decrease in marginal utility to larger domains, by taking multiples of length such as 42 $((\frac{10}{11})^2 = 0.83)$, 63, $((\frac{10}{11})^3 = 0.75)$, 210, $((\frac{10}{11})^{10} = 0.39)$, 2100 $((\frac{10}{11})^{100} = 0.00073)$ and 4200 $((\frac{10}{11})^{200} = 5.26 * 10^{-9})$. Wakker (2010) postulates that such geometric decay is absurd. This implausible result can be accommodated by taking into account reference-dependence (see Bleichrodt et al., 2017, Wakker, 2010).

Additionally, given our aim to apply RP to both individual and societal health outcomes, we will allude to extending this theoretical framework based on Wakker (2010) to a social decision-maker facing gambles over societal health outcomes. We assume that the agent is a social planner deciding from under a Rawlsian veil of ignorance; that is, the agent does not know to whom the outcomes in each gamble will accrue. As we do not allude to (unequal) distributions in our scenarios, for simplicity, we will assume that the societal decision maker aims to maximize health (for a discussions of equity and maximizing health, see: Culyer and Wagstaff, 1993). Furthermore, we will assume monotonicity, i.e. the societal decision maker prefers allocations providing more health for society as a whole over those providing less. Gambles of the form of $g_{0.5}\ell$ indicate a gamble between gaining and losing a small amount of health for society. As in EU, preferences are defined over final population health F_{hp}, which is I_{hp} (initial population health) with the outcomes of gambles incorporated, i.e. the social planners welfare function is reference-independent. Furthermore, we assume that our social planner is risk averse, which, as in EU, is exclusively modelled through the welfare function. When we consider societal outcomes we, thus, assume W to be a concave function over final societal welfare.

Appendix C: On indifference for moderate stakes

The theoretical framework and design used in Chapter 2 are based on strict preferences between rejecting or accepting moderate and large stake gambles. That is, if subjects were indifferent between accepting or rejecting some gamble, this is not captured by our data and does not fit the definitions for RP preferences or preferences that do not violate EU. Typically, indifference is defined as follows: $x \sim y \iff x \succcurlyeq y$ and $x \preccurlyeq y$. Hence, given the choice between x and y, an indifferent subject should be equally likely to pick either of these two outcomes. When we extend this to the stimuli Chapter 2, where subjects were offered the choice to accept or reject gambles of the form $x_{0.5}y$, indifference would imply: $x_{0.5}y \succcurlyeq 0$ (accept) and : $x_{0.5}y \preccurlyeq 0$ (reject). It follows that truly indifferent subjects would be equally likely to accept or reject the gambles, and the results in Table 2.2 give some indication that this may have been the case for individual moderate stake gambles (RP1-RP3). It can be observed that only a slight majority rejected the gamble, which was not significant for RP1. Hence, at least when observing between-subjects, it could be concluded that indifference for moderate stake gambles was likely.

However, if subjects were truly indifferent, it should also hold that rates of acceptance and rejection are split equally *within-subjects*, i.e. a subject who is truly indifferent for any series of gambles should have accepted once or twice, and rejected the other gambles (with fewer exceptions at the extremes of rejecting all three or accepting all three gambles, for which the expected proportion is 12.5%). Table O.2.C.1 shows these proportions for our sample. The large proportions of subjects showing consistent rates of acceptance or rejection throughout, lead to the conclusion that it is unlikely that the majority of the sample was truly indifferent between accepting and rejecting the moderate and large stake gambles in this experiment.

Table O.2.C.1. Proportion of subjects divided by their number (C) of acceptances for gambles within each context. C=0 implies rejection throughout, while C=3 implies acceptance throughout

Gamble	C=0	C=1	C=2	C=3
Individual – moderate stakes	63 (31%)	40 (20%)	85 (42%)	13 (6%)
Individual – large stakes	67 (33%)	96 (48%)	18 (9%)	20 (10%)
Societal – moderate stakes	36 (18%)	36 (18%)	105 (52%)	24 (12%)
Societal – large stakes	32 (16%)	109 (54%)	55 (27%)	5 (2%)

However, it is not possible to rule out that indifference may have occurred. Hence, the definitions of risk aversion, RP preferences, and non-violation of EU would not hold for some subjects. A slight modification of notation and definitions, nonetheless, shows that, for subjects who are in fact indifferent between accepting or rejecting moderate stake gambles, such indifference still implies risk aversion and thus strong concavity on a small domain (as elaborated on in Appendix B of this Chapter). If such preferences, as is assumed in REA, hold for all levels of I_h, or at least on a sufficient range (e.g. in the ranges in Table 0.2.B.1),

indifference for moderate stakes would still imply RP if large stake gambles are accepted. We elaborate below.

Risk aversion is traditionally defined as preferring the expected value of a gamble to (playing) the gamble itself. Hence, it implies $x_{0.5}y \prec 0.5 * x + 0.5 * y$. On the other hand, risk neutrality implies that $x_{0.5}y \sim 0.5 * x + 0.5 * y$. In other words, a risk averse person always prefers the expectation over the gamble while a risk neutral person is indifferent between the expectation and the gamble. By extension, if we assume that $U(\mathcal{H})$ is increasing (which uncontroversially implies that more health gives more utility), a risk neutral person will always have $x_{0.5}y \succ 0$, whenever $0.5 * x + 0.5 * y > 0$, i.e. the gamble has positive expected value. Hence, indifference for moderate stake gambles $g_{0.5}\ell$ with $g = 11$ and ℓ =-10 violates risk neutrality as the expectation of the gamble is positive. Without any further assumptions with regard to the shape of $U(\mathcal{H})$, other than that it is increasing, we can conclude that this person, who is indifferent for moderate stakes gambles, has a certainty equivalent that is smaller than the utility of the expected value (i.e. $U(0)$ > U$(0.5 * g + 0.5 * \ell)$). That is, indifference still implies risk aversion for moderate stakes. It is then straightforward to conclude that if EU with concave $U(\mathcal{H})$ should hold, the analysis in Appendix B of Chapter 2 still holds, except that instead of $U(\mathcal{H})$ dropping by a factor of at least 10/11, it drops by a factor of exactly 10/11, which leads to absurd geometric decay. Hence, indifference for moderate stake gambles implies RP, but the result is weaker.

Given that our definition of RP requires rejection, but indifference is a sufficient condition for RP, this shows that our conclusions are not only robust to this methodological limitation, but it is actually likely that the degree of RP is underestimated. More specifically, subjects that accept moderate stake gambles are currently in no case defined as showing RP preferences, while if they actually were indifferent they may still have violated EU if they accepted a calibrated gamble.

Appendix D: Supplementary analyses and results

Fixed effects model for RP preferences and demographics

We supplement our descriptive findings by running a logistic mixed effects regression models (with the R package lmerTest) on the likelihood of subjects showing RP preferences. The model included subject random effects, and fixed effects for several demographics (as described under 'Additional measures' in Section 3 of the main text): a) age, b) sex, c) BMI, d) subjective health, and e) happiness. Additionally, we include a fixed effect for outcome domain (i.e. individual vs. societal). As can be seen from Table O.2.D.1, these analyses indicate that only outcome domain is a significant predictor of RP preferences, with RP preferences being more likely in the societal domain.

Table O.2.D.1: Results of logistic mixed effects regression predicting the occurrence of RP preferences

	Estimate	*SE*	*z*	*p*
Constant	-1.37	1.07	1.283	0.20
Age	0.006	0.05	0.137	0.89
Sex	0.07	0.12	0.649	0.52
BMI	< -0.001	<0.001	-0.125	0.90
Subjective health	-0.002	0.013	-0.33	0.74
Happiness	0.07	0.022	1.647	0.10
Outcome domain	0.56	0.030	4.82	**<0.001**

Elaboration of results on within-subjects stability of RP preferences (Table 2.3).

Next, in this appendix we report the results of a series of analyses, in which we provide further evidence that the RP preferences observed in this paper a) are not generated by noise, and b) are correlated within scenarios. To that end, we will compare our empirically obtained distribution of RP preferences (Table 2.3) against a series of null-hypotheses, and test if our distribution is significantly different by means of χ^2 tests.

Null hypothesis: all agents satisfy EU

If all agents satisfy EU, no RP preferences should occur, as these preferences by definition violate EU (see section 2 and Appendix B). Thus under this null hypothesis Table 2.3 should have shown the results presented in Table O.2.D.2.

Table O.2.D.2: Frequency (N) and proportion (%) of RP preferences counts (C) within-subjects, if all agents satisfy EU

N (%)		Societal				
		C = 0	C = 1	C = 2	C = 3	Total individual
Individual	C = 0	201 (100%)	0 (0%)	0 (0%)	0 (0%)	201 (100%)
	C = 1	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
	C = 2	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
	C = 3	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
	Total societal	201 (100%)	0 (0%)	0 (0%)	0 (0%)	

Obviously, performing a χ^2–test for this null hypothesis is mathematically impossible, given that no RP preferences were expected at all. That is, by definition under this null hypothesis, when applying a χ^2-testwe will find $\chi^2(df = 15) = \infty, p = 0$ for our observed results in Table 2.3.

Null-hypothesis: all agents act completely randomly or are completely indifferent

If all agent have random preferences the chance that they accept any of the moderate stake- or calibrated gambles is exactly 50%. This also holds if all agents are indifferent between rejecting and accepting both the moderate and large stake gambles. Given that RP preferences are only one out of four possible combinations of preferences (as seen in Table 2.2 in Chapter 2), the chance of showing RP preferences for a single gamble-pair is 25%. Under these assumptions, by probability calculus we can derive the results that Table 2.3 should have given for $n = 201$, as can be seen in Table O.2.D.3.

Table O.2.D.3: Frequency (N) and proportion (%) of RP preferences counts (C) within-subjects, if all agents act completely randomly

N (%)	Societal

		C = 0	C = 1	C = 2	C = 3	Total individual
Individual	C = 0	36 (18%)	36 (18%)	12 (6%)	1 (<1%)	85 (42%)
	C = 1	36 (18%)	36 (18%)	12 (6%)	1 (<1%)	85 (42%)
	C = 2	12 (6%)	12 (6%)	4 (2%)	0 (0%)	28 (14%)
	C = 3	1 (<1%)	1 (<1%)	0 (0%)	0 (0%)	2 (1%)
	Total societal	85 (42%)	85 (42%)	28 (14%)	2 (1%)	

For this null-hypothesis we observed $\chi^2(df = 15) > 15000, p < 0.001$, indicating that if subjects were acting randomly the results we observed in Table 2.3 were extremely unlikely.

Null-hypothesis: all agents are indifferent for moderate stakes gambles

Our design does not allow distinguishing between those that are indifferent between accepting and rejecting gambles or strictly prefer one of these options. If all subjects were indeed indifferent, they would be choosing at random for moderate stakes, and Table 2.2 gives some indication that this may have been the case (only a small majority rejects moderate stake gambles for individual gamble-pairs). Hence, for this null hypothesis, we assumed that all agents are indifferent for moderate stakes (i.e. 50% of accepting or rejecting), and let the probability of accepting or rejecting the large stake gambles depend on the observed frequencies in our sample. Under these assumptions we can derive the expected results for Table 2.3 for *n=201* analytically or by resampling from the empirical distribution (see Table O.2.D.4).

Table O.2.D.4: Frequency (N) and proportion (%) of RP preferences counts (C) within-subjects, if all agents are indifferent for moderate stakes (with empirical distribution informing large stakes).

N (%)		Societal				
		C = 0	C = 1	C = 2	C = 3	Total individual
Individual	C = 0	8 (4%)	16 (8%)	11 (5%)	3 (1%)	38 (19%)
	C = 1	17 (8%)	36 (18%)	26 (13%)	6 (3%)	85 (42%)
	C = 2	13 (6%)	27 (13%)	19 (9%)	4 (2%)	63 (31%)
	C = 3	3 (1%)	7 (3%)	5 (2%)	1 (<1%)	16 (8%)
	Total societal	31 (15%)	86 (43%)	61 (30%)	14 (7%)	

For this null-hypothesis we observed $\chi^2(df = 15) > 946.71, p < 0.001$, indicating that if subjects were truly indifferent for moderate stakes the results we observed in Table 2.3 were extremely unlikely.

Null hypothesis: agents show preferences based on the empirical rates of acceptance and rejection, but these preferences are not correlated between gambles

For this null hypothesis, we assumed that choices on each gamble are drawn independently from the empirical distribution. Consider for example that for RP1 the chance of rejecting the gamble was 109/201 for moderate stakes and 18/201 for large stakes. Under these assumptions we can derive the expected results for Table 2.3 for *n=201* analytically or by resampling from the empirical distribution (see Table O.2.D.5, details and script available on request). This allows us to compare the empirical distribution of RP preferences across individual and societal outcomes to a null hypothesis in which no correlation exists between any gambles.

Table O.2.D.5: Frequency (N) and proportion (%) of RP preferences counts (C) within-subjects, if all agents are indifferent for moderate stakes (with empirical distribution informing large stakes).

N (%)		Societal				
		C = 0	C = 1	C = 2	C = 3	Total individual
Individual	C = 0	2 (1%)	7 (3%)	11 (5%)	6 (2%)	25 (12%)
	C = 1	5 (2%)	22 (11%)	33 (16%)	17 (8%)	77 (38%)
	C = 2	4 (2%)	21 (10%)	32 (16%)	16 (8%)	74 (37%)
	C = 3	1 (<1%)	7 (3%)	10 (5%)	5 (2%)	23 (11%)
	Total societal	13 (6%)	57 (28%)	85 (42%)	43 (21%)	

For this null-hypothesis we observed $\chi^2(df = 15) > 464.27, p < 0.001$, indicating that if all gambles were truly accepted or rejected independently from each other, the results we observed in Table 2.3 were extremely unlikely.

Agents show RP preferences based on the empirical data, but these RP preferences are not correlated between gamble-pairs

For this final null hypothesis, instead of assuming that choices on all gambles are independent of each other, we assumed that each gamble-pair (i.e. a combination of moderate and large stake gamble) is drawn independently from the empirical distribution. That is, for RP1 the chance of showing RP preferences is 94/201, while for RP4 it is 119/201. This procedure allows us to compare the empirical distribution of RP preferences across both outcomes to a null hypothesis where no correlation exists between gamble-pairs for these two outcomes (see Table O.2.D.6).

Table O.2.D.6: Frequency (N) and proportion (%) of RP preferences counts (C) within-subjects, if all agents preferences informed by the empirical distribution (drawn independently).

N (%)		Societal				
		C = 0	C = 1	C = 2	C = 3	Total individual
Individual	C = 0	2 (1%)	10 (3%)	14 (7%)	7 (3%)	25 (12%)
	C = 1	5 (2%)	24 (17%)	35 (16%)	17 (8%)	77 (38%)
	C = 2	5 (2%)	20 (10%)	29 (14%)	14 (8%)	74 (37%)
	C = 3	1 (<1%)	6 (3%)	8 (4%)	4 (2%)	23 (11%)
	Total societal	13 (6%)	60 (30%)	85 (42%)	43 (21%)	

For this null hypothesis we observed $\chi^2(df = 15) > 441.14, p < 0.001$, indicating that if gamble-pairs were truly completed independently across individual and societal tasks, the results we observed in Table 2.3 were extremely unlikely.

Collectively, these five analyses lead to the following conclusion. First, it is impossible that our results were generated by a sample completely consisting of subjects that satisfy EU. Second, it is highly unlikely that our results were generated by a sample consisting completely of subjects that acted randomly, either because of lack of involvement or due to being 'truly' indifferent. Third, our results are not likely to be generated by subjects that were indifferent for moderate stake gambles, as we find more subjects who consistently show RP preferences for all gamble-pairs, or consistently show RP preferences for societal gambles but not for individual gambles. Fourth, these two patterns (i.e. RP throughout and RP for societal but not individual outcomes) occur more frequently than would be expected if no correlation existed between choices on all gambles (Table O.2.D.5) or between gamble-pairs (Table O.2.D.6).

Appendix Chapter 3

This Appendix chapter contains a total of two appendices that provide additional information on the health states used in this experiment (Appendix A), and the experimental instructions and set-up (Appendix B).

Appendix A: Health states used in experiment

See Table O.3.A.1 for more elaborate information on the health states that were utilized in this study.

Table O.3.A.1: Health state descriptions based on EQ-5D-5L

Health state	β0: 11111	β1: 21211	β2: 31221	β3: 32341
Dutch Tariff (Versteegh et al., 2016)	1.00	0.88	0.79	0.46
You have … problems with walking	No	Slight	Moderate	Moderate
You have … problems with washing and dressing yourself	No	No	No	Slight
You have … problems with washing and dressing yourself	No	Slight	Slight	Moderate
… pain or discomfort	No	No	Slight	Severe
… anxious or depressed.	Not	Not	Not	Not

Appendix B: Description of experimental method

Introduction and framing

Subjects were asked to imagine that they would live until 70 years in a health state denoted as health state C. This health state C would be varied for each repetition (4 in total) of the non-parametric method (i.e. β_c). After becoming 70 they were instructed that they would contract a deadly disease, which would lead to a direct, painless death. Their task was to compare two drugs and indicate their preferences between treatments given their health state C and the treatment options, which could be risky, or involve possible side-effects (i.e. losses of life).

Stages of non-parametric method

The non-parametric method is chained, i.e. answers from the previous stage carried over to the next meaning that differences could exist between subjects. For a completely general description of the method we refer to Abdellaoui and colleagues (Abdellaoui et al., 2016). Throughout, as is common for applications of trade-off method (Wakker and Deneffe, 1996), any risky gamble had 50% chance (p = 0.5) of success We denote such gambles as X_pY, meaning X with probability p, and Y otherwise. In our adaptation of the non-parametric method outcomes (i.e. X and Y) reflected life-years. Importantly, throughout it was emphasized that any life years gained or lost were to be spent in health state C. All indifferences were obtained via bi-section. Whenever a variable was elicited it had to be set at some level to start the bi-section method, we chose to obtain this 'starting point' by setting it such that expected value would be equal for both treatments subjects could choose from. For example, if indifference $Z \sim 10_p 0$, elicitation would start at Z=5. This experiment was completely counterbalanced, meaning that health state order and gain-loss order was randomized between subjects. All pre-specified stimuli and elicited indifferences can be found in Table B1.

Stage 1: Coupling gains and losses

Subjects first faced a mixed gamble, which could increase their length of life by G years with probability p, or otherwise decrease it by L years. They could also choose to take a drug that gave 0 years. The negative outcome L was elicited by obtaining the following indifference $G_pL \sim T_0$, where T_0 indicates living until 70 in state C. As can be seen from Table B1, G was fixed at 5, while L was initially set at 2.5 and varied based on individual choices.

Next, two certainty equivalents (CEs) were elicited, which would form the starting points of the standard sequences elicited in Stages 2 and 3. The CE for gains, i.e. the starting point for Stage 2 was elicited by offering subjects a choice between a certain gain x_1^+ in life years (in state C), and a gamble offering G (i.e. 5 years) with probability p, and 0 years otherwise. The amount of life years gained by taking the certain drug (x_1^+) was varied to obtain indifference $x_1^+ \sim G_p T_0$. For losses this procedure was exactly the same, i.e. subjects were offered a choice between a certain drug resulting in a loss of x_1^- life years in state C, and a risky drug. To introduce the loss domain, we instructed them that they had contracted another fatal disease that should also be treated, and thus explained their likely loss compared to T_0 (i.e. 70 years in C). We, thus elicited, $x_1^- \sim L_p T_0$, providing the starting point (x_1^-) for eliciting utility for losses in Stage 3.

Stage 2 & 3: Trade-off method to elicit utility for gains and losses

The trade-off method consists of comparisons between two lotteries. Within our framing, this consisted of two risky drugs, which could increase subjects' life duration in state C to a different extent. Additionally, both drugs could have risks of adverse effects, and thus decrease lifetime in state C. To introduce the loss domain, again subjects were instructed that they had contracted another fatal disease for which treatment was required. Subjects were instructed that they would compare a series of drugs to each other. This series constituted the procedure to elicit the standard sequence, which constitutes of a sequence of outcomes equally spaced in terms of utility (see Wakker and Deneffe, 1996 for proof).

Stage 2, i.e. the trade-off method for gains, commenced by us setting ℓ, a small offset-loss of -1 year in state C. Subjects were offered a choice between two risky drugs: one would offer $x_1^+{}_p\mathcal{L}$, where $\mathcal{L}$ is a larger offset-loss which we aimed to elicit, while the other would offer $\ell_p T_0$. We varied $\mathcal{L}$ to obtain indifference $x_1^+{}_p\mathcal{L}\sim\ell_p T_0$. Next, we elicited standard sequence $(x_2^+, \dots, x_5^+)$ by eliciting indifferences in the form of $x_j^+{}_p\mathcal{L}\sim x_{j-1}^+{}_p\ell$.

Stage 3, i.e. the trade-off method for losses, commenced by us setting $\mathcal{g}$, a small offset-gain of 1 year in state C. Subjects were offered a choice between two risky drugs: one would offer $\mathcal{G}_p x_1^-$ where $\mathcal{G}$ is a larger offset-gain which we aimed to elicit, while the other would offer $\mathcal{g}_p T_0$. We varied $\mathcal{G}$ to obtain indifference $\mathcal{G}_p x_1^-\sim\mathcal{g}_p T_0$. Next, we elicited standard sequence $(x_1^-, x_2^-, \dots, x_5^-)$ by eliciting indifferences in the form of $\mathcal{G}_p x_j^-\sim\mathcal{g}_p x_{j-1}^-$.

Completing this procedure four times for each health state (see Table O.2.A1) resulted in four utility curves, and allowed us to obtain loss aversion parameters and both parametric and non-parametric estimates of utility curvature (see Box 2.I).

Table O.2.B.1: Stimuli used in the three-stage procedure of the non-parametric method.

	Variables elicited	Indifference	Implication	Stimuli
Stage 1	L	$G_p L\sim T_0$	$U(x_1^+) = -U(x_1^-)$	$G = 5\ years$ $p = 0.5$ $T_0 = 70\ years$
	x_1^+	$x_1^+\sim G_p T_0$		
	x_1^-	$x_1^-\sim L_p T_0$		
Stage 2	$\mathcal{L}$	$x_1^+{}_p\mathcal{L}\sim\ell_p T_0$	$U(x_j^+) - U(x_{j-1}^+) = U(x_1^+) - U(0)$	$\ell = -1$ year $j = 5$
	x_j^+	$x_j^+{}_p\mathcal{L}\sim x_{j-1}^+{}_p\ell$		
Stage 3	$\mathcal{G}$	$\mathcal{G}_p x_1^-\sim\mathcal{g}_p T_0$	$U(x_j^-) - U(x_{j-1}^-) = U(x_1^-) - U(0)$	$\mathcal{g} = 1\ year$ $j = 5$
	x_j^-	$\mathcal{G}_p x_j^-\sim\mathcal{g}_p x_{j-1}^-$		

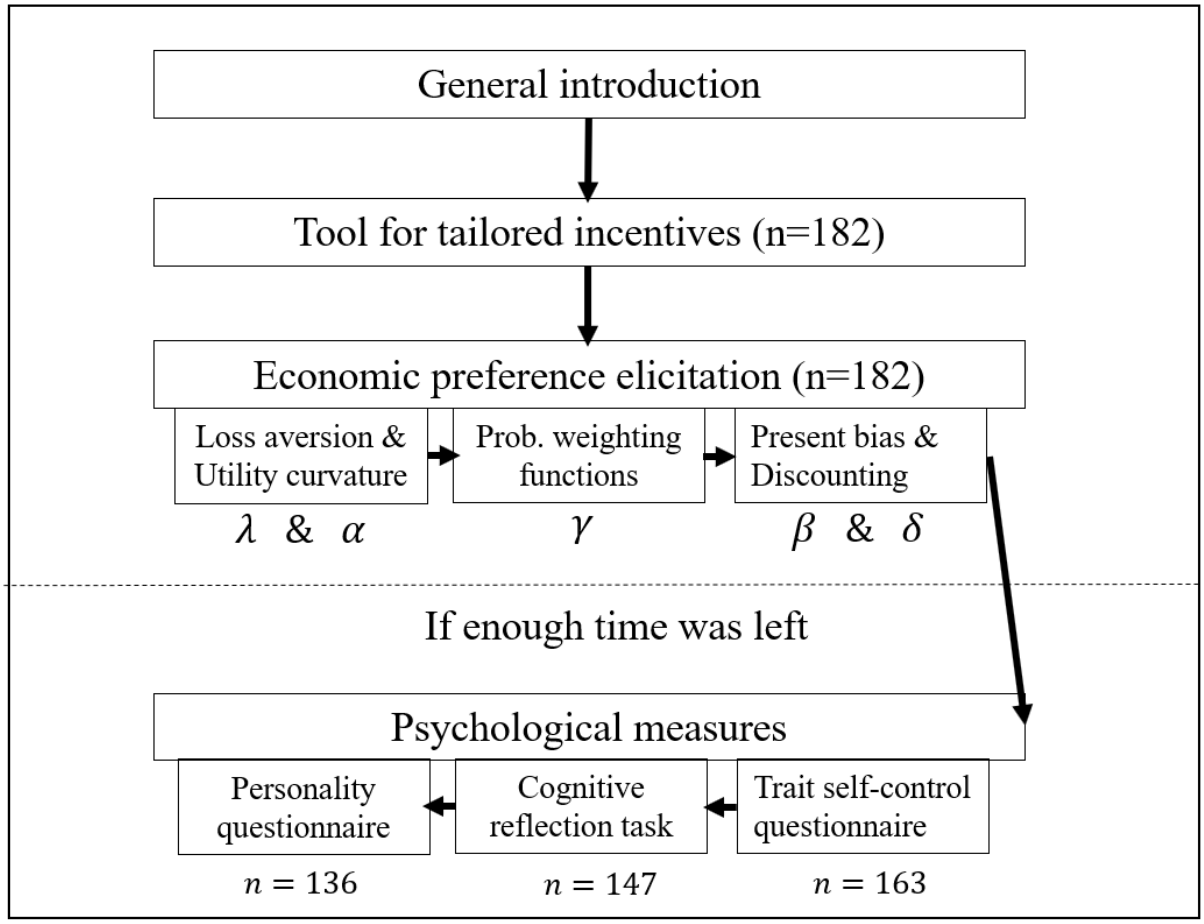
Appendix Chapter 5

This Appendix chapter contains a total of two appendices that provide additional information on the experimental flowchart, instructions, choice task, definitions used to elicit economic preferences. (Appendix A), post-hoc power analysis and interpretation (Appendix B), and supplementary regression results for additional control variables and interactions (Appendix C).

Appendix A: Experimental flowchart, instructions, choice tasks and definitions for economic preferences

Figure O.5.A.1 shows the outline of the experiment, and the measures included in this study. The remainder of this Appendix will provide screenshots and instructions for each part of the experiment.

Figure O.5.A.1. Experimental Flowchart



General introduction

Students were welcomed by the experimenter and received a short oral introduction into the goals of the study. They were told that they could ask any question that they had, and that there were no right or wrong answers.

Tool for tailored incentives

A direct link to this tool can be found here: https://referencepoints.shinyapps.io/Minecentive/

The following instruction was used: *'Please imagine the following situation: you have set yourself the goal of losing weight, so you decided to get a gym membership. Now, your employer wants to help you to lose weight. This may decrease your chances of taking up sick leave and increase your overall wellbeing. As such, your employer has offered to pay you a financial reward if you use the gym at least three times each week for a 10 week period. Your employer is quite flexible, and besides the expected pay-out has no preference in how your*

financial reward is structured. Obviously, you yourself know best what kind of pay-out structure would motivate you to go to the gym and reach your goal of losing weight. Therefore, we ask you to indicate how you would like your pay-out(s) to be structured.'

Next, they could tailor their incentives in a menu, with a separate interactive heading for each of the incentive dimensions. Figures O.5.A.2 to O.5.A5 show the choice options. Subject were given feedback of their selected incentive in the same panel (see Figure O.5.A.6)

Figure O.5.A.2. *Pre-commitment dimension question*

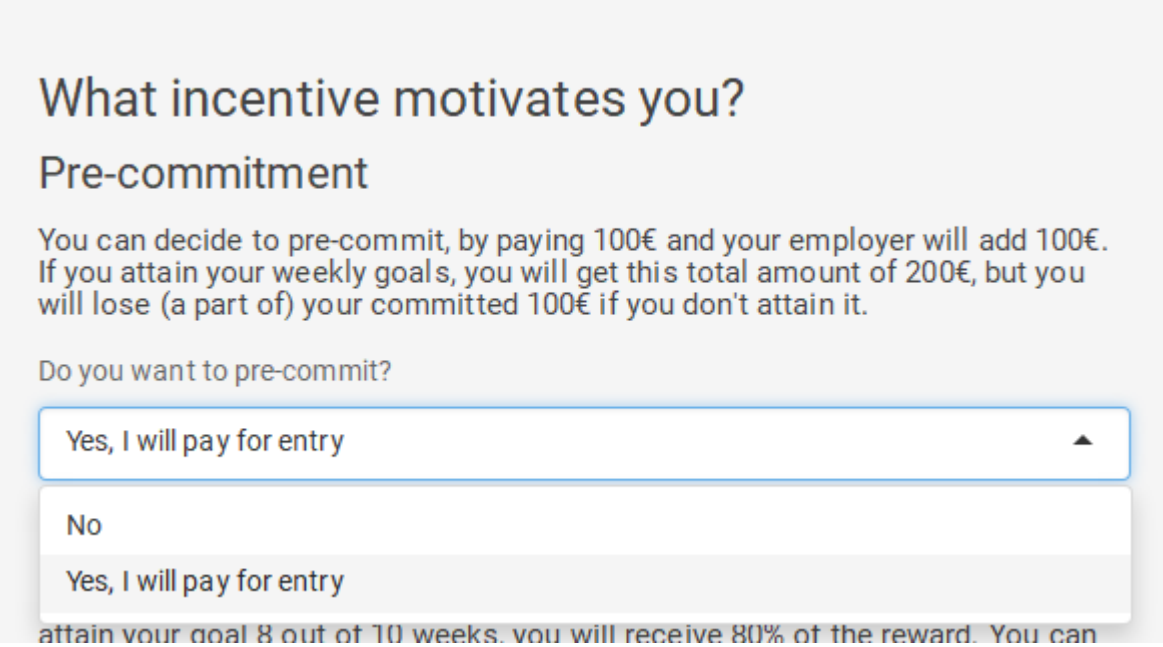


Figure O.5.A.3. *Pay-out frequency*

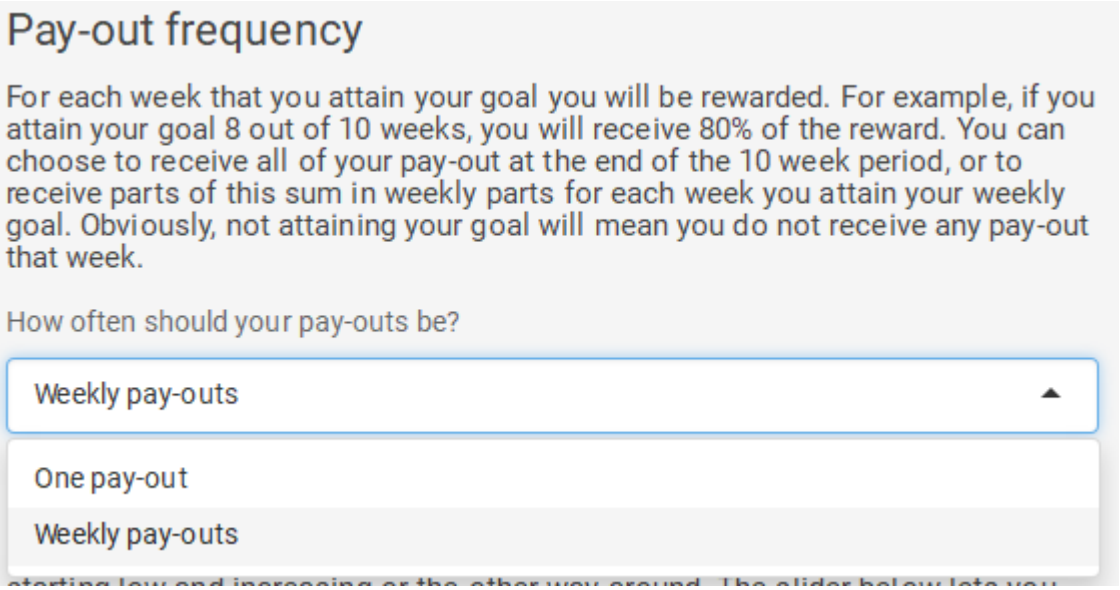


Figure O.5.A.4. *Pay-out structure*

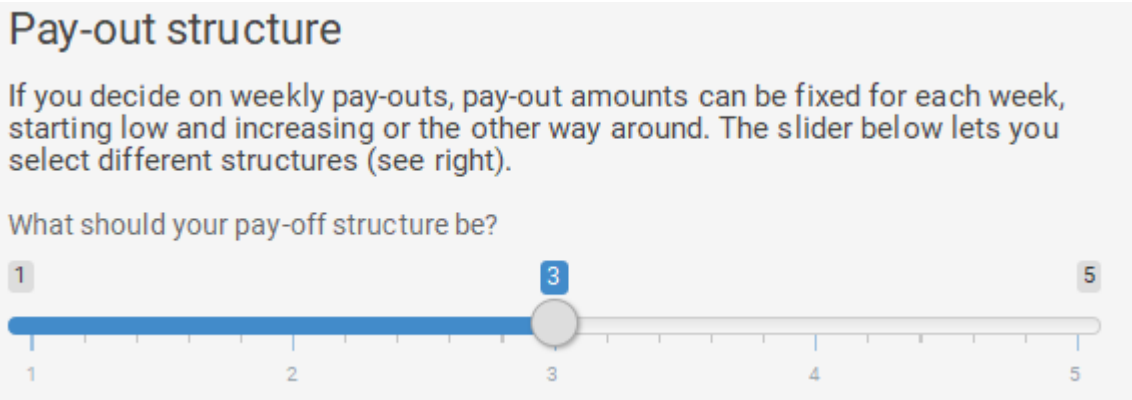


Figure O.5.A.5. *Chance of winning (risk)*

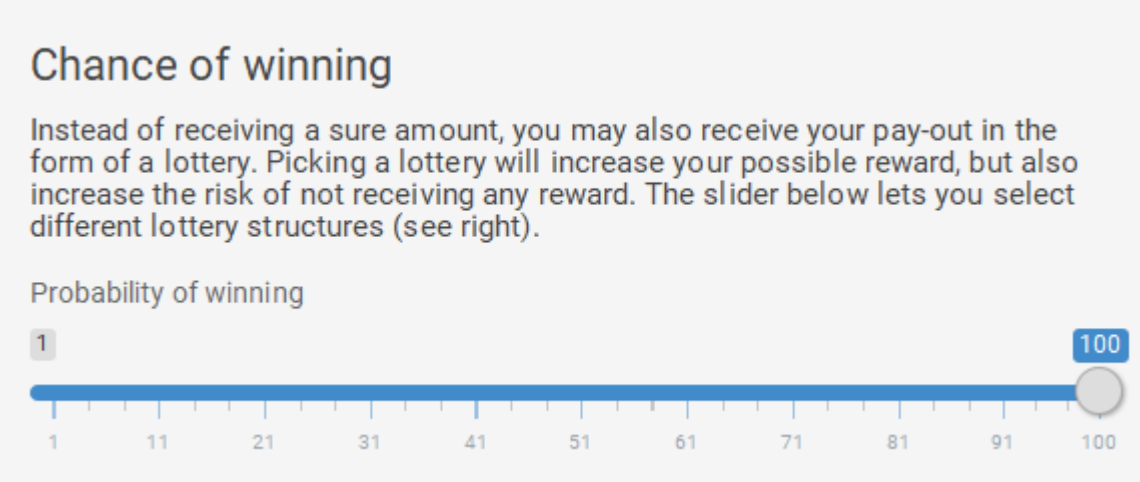


Figure O.5.A.6. *Interactive feedback panel for tailored incentive*

You commit 100 € of your own money

	Reward (Euro)	Chance of winning (%)
Week 1	20	100
Week 2	20	100
Week 3	20	100
Week 4	20	100
Week 5	20	100
Week 6	20	100
Week 7	20	100
Week 8	20	100
Week 9	20	100
Week 10	20	100

Economic preference elicitation

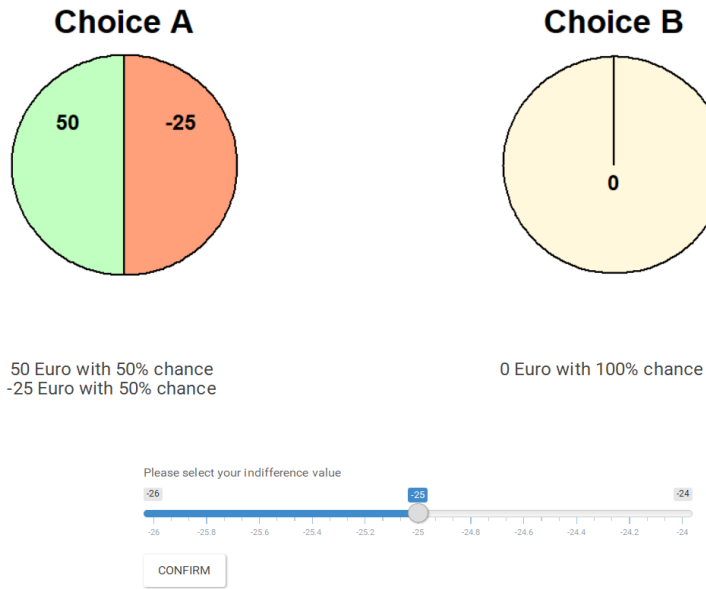
Economic preferences were elicited in three parts in a chained procedure, i.e. responses from each part carried on to the next. The first two parts, aimed at measuring loss aversion, utility curvature and probability weighting were based on the non-parametric method developed by Abdellaoui et al. (2016). A full justification of their approach can be found in the original paper, including the notational conventions, and theoretical assumptions needed to arrive at these elicitations. Throughout, each elicitation consisted of a bi-sectional approach with 4 choices followed by a slider that allowed respondents to modify and confirm their elicited indifference (see Figure O.5.A.7 for an example). Furthermore, throughout this Appendix, all elicitation start with gains first and losses after, while in reality this was counterbalanced between respondents. Throughout this Appendix, we let $\succ, \succcurlyeq, \sim$ represent strict preference, weak preference and indifference respectively.

Loss aversion and utility curvature

Loss aversion

Loss aversion is elicited by eliciting three indifferences, that link gains and losses together. First, we elicit an indifference $g_p l \sim x_0$, where x_0 is the reference-point (set at 0), p is a probability that is kept constant throughout this first, and g is a gain of €50. We elicit a loss l (e.g. -25€). The next two indifferences involve certainty equivalence elicitation for outcome g and l, i.e. a certain outcome (x_1^+ for gains, x_1^- for losses) that makes one indifferent between receiving g or l with probability p or x_0 otherwise. These indifferences are denoted $x_1^+ \sim g_p x_0$ and $x_1^- \sim l_p x_0$. Abdellaoui et al. (2016) show that loss aversion (denoted as index λ), as defined by Köbberling and Wakker (2005), can be derived by: $\lambda = x_1^+ / -x_1^-$, where respondents with $\lambda > 1$, $\lambda = 1, \lambda < 1$ are loss averse, loss neutral or gain seeking respectively.

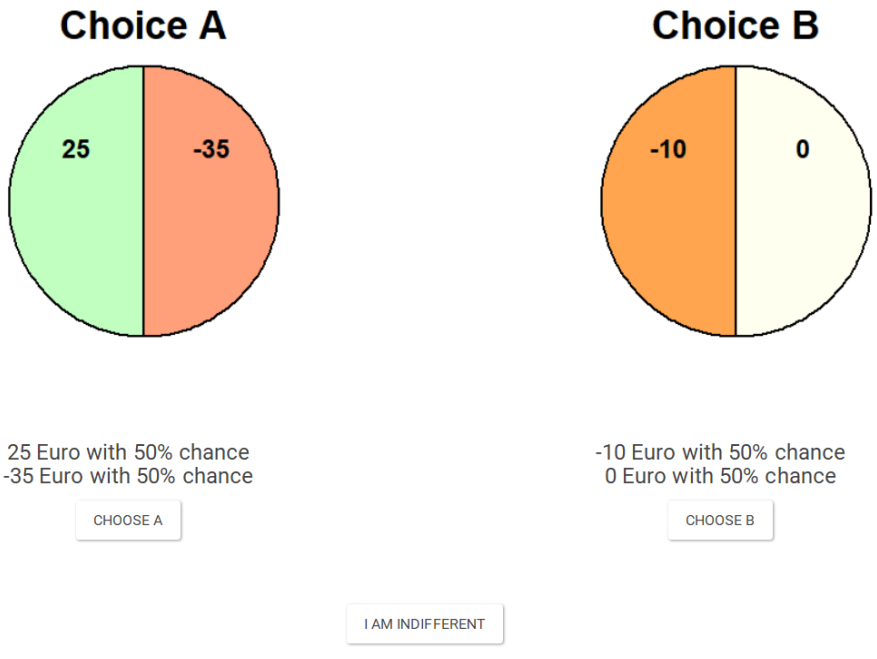
Figure O.5.A.7. Example of elicitation procedure for indifference $g_p l \sim x_0$, where a respondent is indifferent for $l = 25$



Utility curvature

Now that we have linked together gains and losses compared to the reference-point x_0, we can elicit a series of indifference to estimate utility curvature for both gains and losses separately. In Abdellaoui et al. (2016) this process is based on the trade-off method developed by Wakker and Deneffe (1996), see Figure O.5.A.8. That is, a standard sequence of outcomes is elicited for gains and losses, or in other words a sequence of outcomes spaced equally in terms of utility. This standard sequence elicitation is set-up in the same way for gains and losses. For gains, it starts by fixing a small loss ℓ (in this study fixed at: -10€), and eliciting a larger loss $\mathcal{L}$ in the following indifference: $x_1^+ {}_p\mathcal{L} \sim \ell_p x_0$. These two loss amounts serve as offset losses, in the standard sequence elicitation for gains. Next, the equally-spaced outcomes in the standard sequence are elicited by eliciting x_2^+ in the following indifference : $x_2^+ {}_p\mathcal{L} \sim \ell_p x_1^+$. This process (i.e. $x_j^+ {}_p\mathcal{L} \sim x_{j-1}^+ {}_p\ell, j = 2, \dots 4$) is applied 3 times, yielding a standard sequence with 5 data points (x_0, x_1^+, x_2^+, x_3^+, and x_4^+). For losses, a small gain (in this study fixed at: 10€) is fixed, to elicit a larger loss $\mathcal{G}$, in the following indifference: $\mathcal{G}_p x_1^- \sim g_p x_0$. Next, again a series of indifferences of the form $\mathcal{G}_p x_j^- \sim g_p x_{j-1}^-, j = 2, \dots, 4$ is elicited, which yields a standard sequence for losses with 5 data points (x_0, x_1^-, x_2^-, x_3^-, and x_4^-).

Figure O.5.A.8. Example visual representation of elicitation procedure for utility curvature (indifference $x_1^+ {}_p\mathcal{L} \sim \ell_p x_0, with\ x_1^+ = 25$)



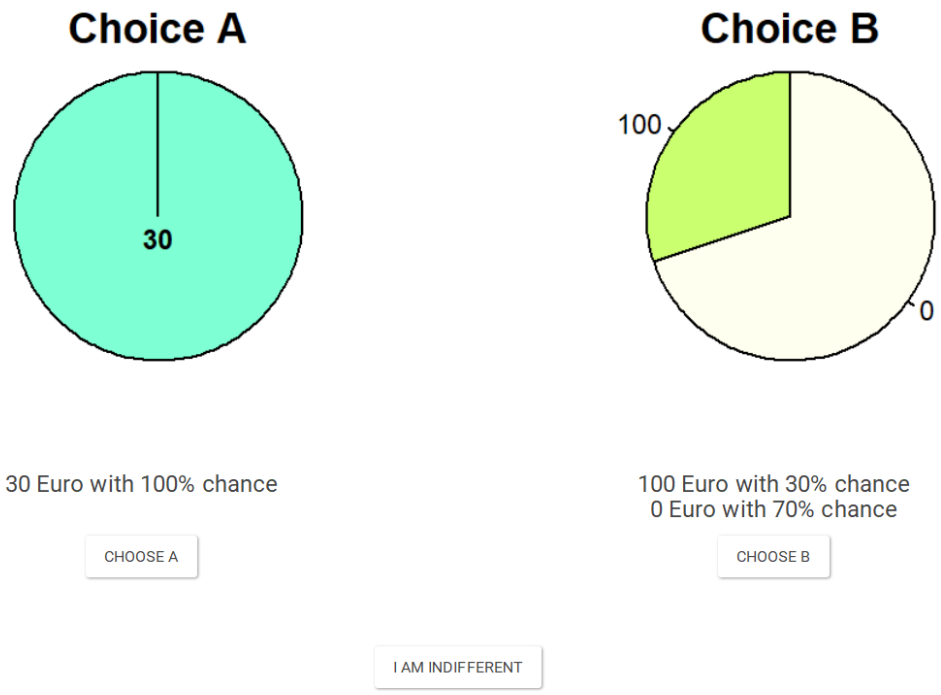
To calculate the utility curvature of the utility function for monetary gains $U^+(\cdot)$ or losses $U^-(\cdot)$, we apply the following scaling: $U^+(x_0) = 0, U^-(x_4^-) = -1$ and $U^+(x_4^-) = 1$. Furthermore, as is usual in these type of studies, monetary outcomes are normalized such that each outcome is dived by the highest outcome in its' respective domain, i.e. x_0, x_1^-, x_2^-, x_3^-, x_4^- / x_4^- and $x_0, x_1^+, x_2^+, x_3^+, x_4^+ / x_4^+$. Although the non-parametric method allows non-parametric estimation of utility curvature, in this study the most commonly used power utility family is used estimated by non-linear least squares. This allows the estimates to be

compared with earlier work. For this family, $U^{+}(x) = x^{\alpha}$, where represents the utility function over monetary outcomes. For losses, this is estimated by $U^{-}(x) = -(-(x)^{\alpha}$ with $\alpha > 0$. For gains [losses], $\alpha > 1$ corresponds to convex [concave] utility, $\alpha = 1$ corresponds to linear utility, and $\alpha < 1$ corresponds to concave [convex] utility.

Probability weighting functions

As in Chapter 7 of this dissertation, probability weighting is elicited with the method developed by Abdellaoui (2000). This method was employed as follows: to the probability weighting functions $w^{+}(p)$ and $w^{-}(p)$, the certainty equivalents x_p^{+} and x_p^{-} of the prospects $x_{4_p}^{+}x_0$ and $x_{4_p}^{-}x_0$, for the following probabilities: $p = 0.1, 0.3, 0.5, 0.7, 0.9$.The outcomes x_4^{+} and x_4^{-} are the maximum (minimum) outcome elicited in the standard sequence. Therefore, it follows from the chosen scaling of utility that $U(x_P^{+}) = w^{+}(p)$ and $-U^{-}(x_p^{-}) = w^{-}(p)$. The values of $U^{+}(x_p^{+})$ and $LU^{-}(x_p^{-})$ are interpolated from their respective standard sequences. Figure O.5.A.9. shows an example of a gamble scenario for gains.

Figure O.5.A.9. Example visual representation of choice options used for eliciting probability weighting for gains $(x_p^{+} \sim x_{4_p}^{+}x_0), with\ x_4^{+} = 100, p = 0.3$.



To summarize the shape of the weighting functions Tversky and Kahneman's one-parameter inverse S-shaped probability weighting function is used, i.e. $w^{i}(p) = p^{\gamma}/(p^{\gamma} + (1-p)^{\gamma})^{1/\gamma}$ with $i = +, -$. Again, this is estimated by nonlinear least squares. The γ-parameter controls for the shape of the probability weighting function. If $\gamma = 1$ there is no probability transformation and $w^{i}(p) = p$. However, if $\gamma < 1$, decision makers underweight large probabilities and overweight small probabilities. This corresponds to the commonly found inverse S-shaped weighting function. If $\gamma > 1$, the opposite pattern holds, corresponding to an S-shaped weighting function.

Present bias & Discounting

Present bias and discounting were elicited by means of the approach of Laibson (1997). This model assumes the discounted utility model, i.e. utility for timed outcome (x, t) can be evaluated by $DU(x,t) = D(t)U(x)$, where $D(\cdot)$ refers to the discounting function. To reflect the sign-dependent nature of this experiment, we modify this to: $U(x,t) = D^i(t)U^i(x)$, with $i = +$ for gains and $i = -$ for losses. In the quasi-hyperbolic discounting model, $D^i(t) = \beta^i/(1+r^i)^t$, with $i = +, -$ for gains and losses respectively, $0 < \beta \leq 1$ for $t > 0$ and $D(t) = 1$ otherwise, and r reflects the per-period discount rate. In this part of the elicitation x_T^+ and x_T^- refer to the highest outcome in the standard sequences for gains and losses divided by 2, i.e. $x_4^+/2$ and $x_4^-/2$. These outcomes were divided by half to decrease the chances extrapolation beyond the measured standard sequence was necessary. β and r were elicited for gains and losses by means of the following indifferences (see Figure O.5.A.10 for an example of visual stimuli used),

$$(x_T^+, 0\ weeks) \sim (y_T^+, 5\ weeks)$$

$$(x_T^+, 5\ weeks) \sim (z_T^+, 10\ weeks)$$

And for losses by:

$$(x_T^-, 0\ weeks) \sim (y_T^-, 5\ weeks)$$

$$(x_T^-, 5\ weeks) \sim (z_T^-, 10\ weeks)$$

Figure O.5.A.10. Visual representation of choice options used for eliciting present bias and discounting for losses.

In both cases, we can evaluate these indifferences as:

$$U^i(x_T^i)D^i(0) = U^i(y_T^i)D^i(5) \iff U^i(x_T^i)\left(\frac{1}{(1+r^i)^1}\right) = U^i(y_T^i)\left(\frac{\beta^i}{(1+r^i)^5}\right)$$

$$U^i\left(x_T^i\right)D^i(5) = U^i(z_T^i)D^i(10) \iff U^i(x_T^i)\left(\frac{\beta^i}{(1+r^i)^5}\right) = U^i(y_T^i)\left(\frac{\beta^i}{(1+r^i)^{10}}\right)$$

After rearranging the second indifference we find:

$$r^i = \frac{1}{\left(\frac{U^i(z_T^i)}{U^i(x_T^i)}\right)^{1/5}}$$

And after we have determined r^i, β is found by:

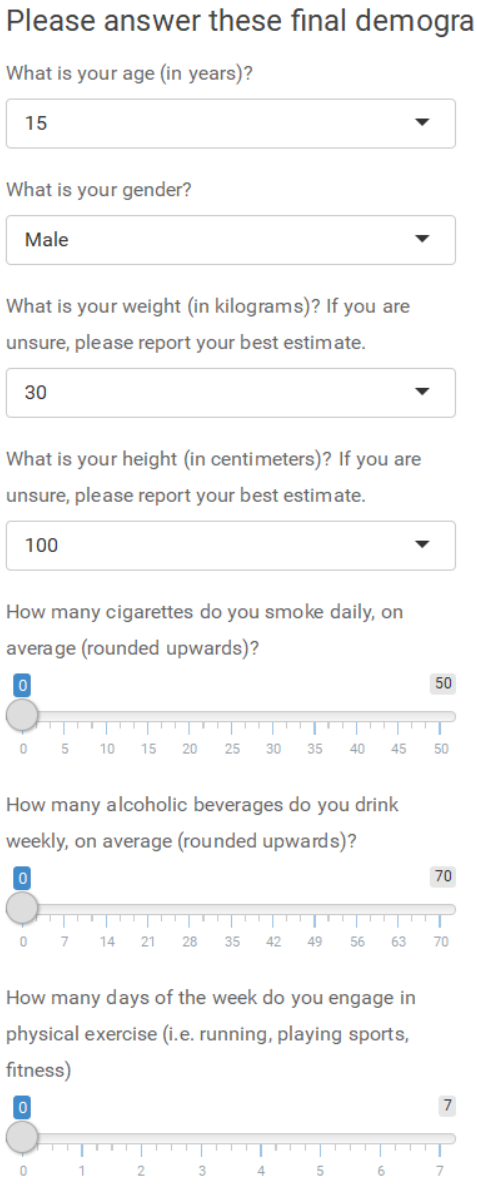
$$\beta^i = \frac{U^i(y_T^i)}{U^i(x_T^i)\frac{1}{(1+r^i)^{1/5}}}$$

Psychological measures

After asking respondents to self-report on several health-related characteristics, total of three psychological measures were used, which are reprinted in this Appendix. These questionnaires measured self-control (Tangney et al., 2018), cognitive reflection (Toplak et al., 2011), and personality (Francis et al., 1992).

Self-reported health behavior

The following demographics were collected:



Trait self-control questionnaire

The questionnaire was adapted from Tangney et al. (2018), and measures self-control as a trait, i.e. the degree to which individuals in general are able to self-regulate. Items marked with * require reverse coding, and it is reported as a mean in the main text.

The following statements may reflect how you perceive yourself. Please indicate below to what extent these statements reflect how you typically are, by circling the answer that applies.

		Not at all ↓				***Very much*** ↓
1	*I am good at resisting temptation*	*1*	*2*	*3*	*4*	*5*
*2**	*I have a hard time breaking bad habits.*	*1*	*2*	*3*	*4*	*5*
*3**	*I am lazy*	*1*	*2*	*3*	*4*	*5*
*4**	*I say inappropriate things.*	*1*	*2*	*3*	*4*	*5*
*5**	*I do certain things that are bad for me, if they are fun.*	*1*	*2*	*3*	*4*	*5*
6	*I refuse things that are bad for me.*	*1*	*2*	*3*	*4*	*5*
*7**	*I wish I had more self-discipline.*	*1*	*2*	*3*	*4*	*5*
8	*People would say that I have iron self- discipline.*	*1*	*2*	*3*	*4*	*5*
*9**	*Pleasure and fun sometimes keep me from getting work done.*	*1*	*2*	*3*	*4*	*5*
*10**	*I have trouble concentrating.*	*1*	*2*	*3*	*4*	*5*
*11**	*I am able to work effectively toward long-term goals.*	*1*	*2*	*3*	*4*	*5*
*12**	*Sometimes I can't stop myself from doing something, even if I know it is wrong.*	*1*	*2*	*3*	*4*	*5*
*13**	*I often act without thinking through all the alternatives.*	*1*	*2*	*3*	*4*	*5*

Cognitive reflection task (CRT)

This three item task developed by Toplak et al. (2011) aims to quantify the degree to which individuals rely on their automatic system by asking questions which seems to have an immediate, simple and right answer, which only after reflecting on it for some time appears to be in fact *wrong.* The cognitive reflection task is scored as the amount of correct answers. The questions were answered by with a pen by writing down the answer on the open space.

1. *A bat and a ball cost €1.10 in total. The bat costs €1.00 more than the ball. How much does the ball cost? _____ cents*
2. *If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? _____ minutes*
3. *In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take*

Personality questionnaire

The last questionnaire used is a revised short-form version of the Revised Eysenck Personality Questionnaire, which captures personality on 4 domains. Items 1, 10, 12, 15, 19 and 22 capture Neuroticism. Items 2, 4, 14, 16, 21 and 24 capture Extraversion. Items 3, 6, 9, 13, 17 and 23 capture Psychoticism, and finally, items 5, 7, 8, 11, 18, and 20 capture Social desirability. Items marked with * are recoded, meaning that code 1 means has characteristic related to personality dimension and code 0 means does not relate to that dimension. Means are reported in main text.

Please answer the following questions by answering "Yes" or "No" (circle which applies). There are no right or wrong answers. It is not necessary to think very long about these questions.

1	*Does your mood often go up and down?*	*Yes*	*No*
2	*Are you a talkative person?*	*Yes*	*No*
*3**	*Would being in debt worry you?*	*Yes*	*No*
4	*Are you rather lively?*	*Yes*	*No*
*5**	*Were you ever greedy by helping yourself to more than your share of anything?*	*Yes*	*No*
6	*Would you take drugs which may have strange or dangerous effects?*	*Yes*	*No*
*7**	*Have you ever blamed someone for doing something you knew was really your fault?*	*Yes*	*No*
8	*Do you always practice what you preach?*	*Yes*	*No*
9	*Do you prefer to go your own way rather than act by the rules?*	*Yes*	*No*
10	*Do you often feel 'fed-up'?*	*Yes*	*No*
*11**	*Have you ever taken anything (even a pin or button) that belonged to someone else?*	*Yes*	*No*
12	*Would you call yourself a nervous person?*	*Yes*	*No*
13	*Do you think marriage is old-fashioned and should be done away with?*	*Yes*	*No*
14	*Can you easily get some life into a rather dull party?*	*Yes*	*No*
15	*Are you a worrier?*	*Yes*	*No*
*16**	*Do you tend to keep in the background on social occasions?*	*Yes*	*No*
*17**	*Does it worry you if you know there are mistakes in your work?*	*Yes*	*No*
*18**	*Have you ever cheated at a game?*	*Yes*	*No*
19	*Do you suffer from 'nerves'?*	*Yes*	*No*
*20**	*Have you ever taken advantage of someone?*	*Yes*	*No*
*21**	*Are you mostly quiet when you are with other people?*	*Yes*	*No*
22	*Do you often feel lonely?*	*Yes*	*No*
*23**	*Is it better to follow society's rules than go your own way?*	*Yes*	*No*
24	*Do other people think of you as being very lively*	*Yes*	*No*

Appendix B. Post-hoc power analysis and interpretation of effect sizes

This Appendix will reflect on the statistical power of the experiment reported. Seeing as this is the first study that aims to study the association between tailored incentives and behavioral insights (obtained as stated preferences), no effect sizes were available for a-priori power analysis. Furthermore, a-priori power analyses are often performed for studies with random assignment, which ensures that samples are independent and gives the researcher control over the number of respondents in each group. In this study, however, it is questionable if samples are independent, and barely any work existed to predict how individuals self-select incentives. Hence, in my view, only post-hoc power analyses were available, which I report in this appendix.

For simplicity, I will base power analysis on independent samples *T* tests with 2 samples, as this will easily allow me to demonstrate the lowest mean differences this study was powered to detect. This means that for each incentive dimension I divide the self-selected incentives into 2 samples, that I treat as being independent. The lack of significant results in my study indicates that for most measures my mean differences were lower. If this study is underpowered, it runs the risk of unjustly categorizing these small differences as error rather resulting from distributions with a true mean difference. Any study has such a risk of such a Type II error, which is captured by (1-β), where β reflects the desired test power, which I set at 0.8, as is suggested by Cohen (1988). However, it is impossible to determine if such an error actually occurred. Hence, whether or not this study is sufficiently powered is a value judgment to be made by reflecting on the economic significance of the lowest mean differences this study is able to detect. If one believes that mean differences smaller than this study was able to detect are economically significant (i.e. are meaningful in terms of stated preferences or real decisions), this study is underpowered. If on the other hand, one believes mean differences of the order of magnitude reported in Table O.5.B.1 to Table O.5.B.4 to be negligible, this study was adequately powered.

First, the sample and its' empirical distribution over the tailored incentive dimensions is used to determine the lowest effect size this study is powered to observe (Table O.5.B.1-4, *Cohen's d*). For Pre-commitment and Weekly, which had two answer categories, this will be based on an independent samples *t* test. The minimal effect size is a function of the size of each group (Table O.5.B.1-4, *n1* and *n2*), significance level ($\propto = 0.05$), and the preferred statistical power (β=0.8). This approach can also be applied to Structure and Risk, but this first requires dichotomizing these dimensions. Structure is dichotomized as those selecting those who chose constant (32%) or one of the non-constant payment structures (68%). Risk will be dichotomized as those who chose certain pay-outs (48%) and those who introduced some risk (52%).

Cohen's d is calculated as follows: $d = M2 - M1/(SD_{pooled})$., i.e. the mean difference divided by the pooled standard deviation. Hence, for each observation, I calculated the pooled standard deviation, and multiply by Cohen's *d* to obtain the smallest parameter difference this study was powered to observe. Table O.5.B.1 to O.5.B.4 report the outcomes of this analysis (and the inputs used). It shows that Cohen's d ranged from 0.42 to 0.51, which are between small and

medium effect sizes using Cohen's (1988) interpretation. Whether or not the lowest mean difference this study was powered to observe includes economically significant differences is ultimately a value judgement. My conclusion is that due to the high heterogeneity observed for the economic preferences elicited, the study runs some risk of being underpowered. For example, for effects of discounting for gains and losses the lowest mean differences this study was powered to observe was 0.05 and 0.03 respectively. Given that these are *weekly* discount rates, differences much smaller than these could very well be economically significant. However, given the exploratory nature of this study, many measures were collected, and none appeared to significantly and systematically predict tailored incentives. Hence, it is straightforward to see that it is unlikely that all of these null-results (i.e. the small actual differences reported in Table O.5.B.1-4) are Type II errors. In fact, without correcting for multiple testing it would have been far more likely to have found significant but non-existing effects (i.e. Type I errors).

Table O.5.B.1. Post-hoc power analysis for Pre-commit dimension of tailored incentives

Dimension: Pre-commitment								
	N1	***N2***	**Cohen's *d***	***SD1***	***SD2***	SD_{pooled}	**Lowest difference**	**Actual difference**
Loss aversion (λ)	59	123	0.45	3.15	2.22	2.73	1.22	0.65
Utility curvature (α) - gains	59	123	0.45	1.53	3.07	2.43	1.08	-0.24
Utility curvature (α) - losses	59	123	0.45	4.72	3.93	4.34	1.94	0.29
Probability weighting (γ) - gains	59	123	0.45	2.67	2.05	2.38	1.06	0.05
Probability weighting (γ) - losses	59	123	0.45	2.89	2.32	2.62	1.17	0.51
Present Bias (β) - gains	58	123	0.45	0.18	0.33	0.27	0.12	-0.05
Present Bias (β) - losses	59	123	0.45	0.18	0.21	0.2	0.09	0
Discounting (δ) - gains	59	123	0.45	0.05	0.14	0.1	0.05	-0.02
Discounting (δ) - losses	59	123	0.45	0.04	0.08	0.07	0.03	-0.02
Age	59	123	0.45	1.45	1.49	1.47	0.66	0
Cigarettes (per week)	59	123	0.45	2.82	2.57	2.7	1.2	0.04
BMI	59	123	0.45	5.98	3.13	4.78	2.13	-0.35
Alcohol (glasses/week)	59	123	0.45	8.34	9.93	9.17	4.09	-0.35
Exercise (days/week)	59	123	0.45	1.74	1.64	1.69	0.75	-0.19
Trait self-control	55	107	0.47	0.59	0.57	0.58	0.27	0.03
Cognitive reflection	50	96	0.49	1.16	1.16	1.16	0.57	0.13
EPQ - Neuroticism	45	90	0.52	0.22	0.18	0.2	0.11	-0.06
EPQ - Extraversion	45	90	0.52	0.19	0.17	0.18	0.09	-0.01
EPQ - Psychoticism	45	90	0.52	0.2	0.18	0.19	0.1	-0.04
EPQ – Social desirability	45	90	0.52	0.23	0.22	0.23	0.12	0.05

Table O.5.B.2. Post-hoc power analysis for Weekly dimension of tailored incentives

Dimension: Weekly								
	N1	***N2***	**Cohen's *d***	***SD1***	***SD2***	SD_{pooled}	**Lowest difference**	**Actual difference**
Loss aversion (λ)	86	96	0.42	2.67	2.49	2.58	1.08	0.08
Utility curvature (α) - gains	86	96	0.42	3.79	0.69	2.73	1.14	0.7
Utility curvature (α) - losses	86	96	0.42	0.81	5.7	4.07	1.7	-0.88
Probability weighting (γ) - gains	86	96	0.42	2.86	1.52	2.29	0.96	0.47
Probability weighting (γ) - losses	86	96	0.42	1.61	3.09	2.47	1.03	-0.67
Present Bias (β) - gains	86	95	0.42	0.34	0.24	0.3	0.12	0.04
Present Bias (β) - losses	86	96	0.42	0.23	0.18	0.2	0.08	-0.06
Discounting (δ) - gains	86	96	0.42	0.13	0.1	0.12	0.05	0.02
Discounting (δ) - losses	86	96	0.42	0.06	0.09	0.07	0.03	0
Age	86	96	0.42	1.38	1.56	1.47	0.62	-0.04
Cigarettes (per week)	86	96	0.42	2.73	2.57	2.65	1.11	0.39
BMI	86	96	0.42	5.3	2.98	4.3	1.8	1.08
Alcohol (glasses/week)	86	96	0.42	10.03	8.89	9.48	3.96	0.58
Exercise (days/week)	86	96	0.42	1.64	1.69	1.67	0.7	-0.19
Trait self-control	81	81	0.44	0.56	0.59	0.57	0.25	-0.13
Cognitive reflection	74	72	0.47	1.12	1.17	1.15	0.54	0.33
EPQ - Neuroticism	70	65	0.49	0.21	0.18	0.2	0.1	-0.05
EPQ - Extraversion	70	65	0.49	0.18	0.18	0.18	0.09	0
EPQ - Psychoticism	70	65	0.49	0.18	0.19	0.19	0.09	-0.03
EPQ – Social desirability	70	65	0.49	0.22	0.24	0.23	0.11	0.01

Table O.5.B.3. Post-hoc power analysis for Weekly dimension of tailored incentives

Dimension: Sequence								
	N1	***N2***	**Cohen's *d***	***SD1***	***SD2***	SD_{pooled}	**Lowest difference**	**Actual difference**
Loss aversion (λ)	117	65	0.44	3.05	1.3	2.35	1.02	0.47
Utility curvature (α) - gains	117	65	0.44	3.28	0.74	2.37	1.03	0.51
Utility curvature (α) - losses	117	65	0.44	5.16	1.06	3.73	1.62	0.5
Probability weighting (γ) - gains	117	65	0.44	2.53	1.69	2.15	0.94	0.24
Probability weighting (γ) - losses	117	65	0.44	2.03	3.2	2.68	1.17	-0.64
Present Bias (β) - gains	117	64	0.44	0.32	0.25	0.28	0.12	0
Present Bias (β) - losses	117	65	0.44	0.2	0.2	0.2	0.09	-0.04
Discounting (δ) - gains	117	65	0.44	0.14	0.06	0.11	0.05	0.03
Discounting (δ) - losses	117	65	0.44	0.05	0.1	0.08	0.03	-0.01
Age	117	65	0.44	1.41	1.6	1.51	0.66	0.1
Cigarettes (per week)	117	65	0.44	2.8	2.37	2.59	1.13	0.26
BMI	117	65	0.44	4.65	3.37	4.06	1.77	1.04
Alcohol (glasses/week)	117	65	0.44	9.91	8.55	9.26	4.03	0.28
Exercise (days/week)	117	65	0.44	1.63	1.75	1.69	0.74	-0.2
Trait self-control	108	54	0.47	0.57	0.59	0.58	0.27	-0.03
Cognitive reflection	97	49	0.49	1.14	1.17	1.16	0.57	0.28
EPQ - Neuroticism	91	44	0.52	0.21	0.17	0.19	0.1	-0.05
EPQ - Extraversion	91	44	0.52	0.18	0.18	0.18	0.09	-0.03
EPQ - Psychoticism	91	44	0.52	0.18	0.19	0.19	0.1	-0.06
EPQ – Social desirability	91	44	0.52	0.22	0.24	0.23	0.12	-0.02

Table O.5.B.4. Post-hoc power analysis for Weekly dimension of tailored incentives

Dimension: Risk								
	N1	***N2***	**Cohen's *d***	***SD1***	***SD2***	SD_{pooled}	**Lowest difference**	**Actual difference**
Loss aversion (λ)	87	95	0.42	2.6	2.56	2.58	1.08	0.09
Utility curvature (α) - gains	87	95	0.42	2.67	2.68	2.68	1.12	0.03
Utility curvature (α) - losses	87	95	0.42	3.93	4.44	4.19	1.75	0.01
Probability weighting (γ) - gains	87	95	0.42	2.47	2.05	2.27	0.95	0.35
Probability weighting (γ) - losses	87	95	0.42	3.11	1.81	2.54	1.06	0.54
Present Bias (β) - gains	87	94	0.42	0.21	0.35	0.29	0.12	-0.04
Present Bias (β) - losses	87	95	0.42	0.17	0.23	0.2	0.08	0.05
Discounting (δ) - gains	87	95	0.42	0.12	0.11	0.12	0.05	0
Discounting (δ) - losses	87	95	0.42	0.08	0.06	0.07	0.03	-0.01
Age	87	95	0.42	1.57	1.38	1.48	0.62	0.25
Cigarettes (per week)	87	95	0.42	2.41	2.85	2.64	1.1	-0.31
BMI	87	95	0.42	5.07	3.38	4.3	1.8	0.09
Alcohol (glasses/week)	87	95	0.42	9.58	9.33	9.46	3.95	-0.35
Exercise (days/week)	87	95	0.42	1.59	1.74	1.67	0.7	0.08
Trait self-control	79	83	0.44	0.57	0.59	0.58	0.26	-0.07
Cognitive reflection	72	74	0.47	1.15	1.17	1.16	0.54	-0.05
EPQ - Neuroticism	69	66	0.49	0.18	0.21	0.2	0.1	0.01
EPQ - Extraversion	69	66	0.49	0.17	0.19	0.18	0.09	0
EPQ - Psychoticism	69	66	0.49	0.2	0.18	0.19	0.09	0.02
EPQ – Social desirability	69	66	0.49	0.21	0.24	0.23	0.11	-0.03

Appendix C: Regression results including additional control variables and interactions

A full overview of all models ran can be found below (Table O.5.C.1, which confirm that selected incentives could not reliably be predicted from any of the measures collected (except BMI for the timing dimension). For the Timing dimension, after many exploratory regression analyses, a model with some significant predictors could be developed, which is reported in Table O.5.C.2. Due to the exploratory process through which these results were obtained, no conclusions are based on it in the main text. Although more model specifications were possible, any correction for multiple hypothesis testing (which would be advised given the plethora of tests applied here) would quickly lead to null results.

Table O.5.C.1: All models ran, including significant ($p < 0.05$) predictors (**boldfaced**), adjusted R-squared, Akaike's Information Criterions (AIC) and Bayesian Information Criterion (BIC).

Note: All models are specified as R model formulas, where x~y indicates predicting x by y. * signifies that these economic preferences were also not a significant predictor of incentive choice after controlling for all demographics and/or psychological measures,

Outcome	**Model ran**	R^2	**AIC**	**BIC**
Pre-commit (PRC)	Logistic regression			
	Economic preferences			
	PRC ~ Loss aversion*	0.012	231.15	233.13
	PRC ~ Utility curvature (losses)*	<0.001	233.13	239.54
	PRC ~ Probability weighting (losses)*	0.009	231.75	238.2
	PRC ~ Present bias (losses)*	0.004	233.24	239.65
	PRC ~ Discounting (losses)*	0.009	231.37	237.78
	PRC ~ Loss aversion + Utility curvature (losses) + Probability weighting (losses) + Present bias (losses) + Discounting (losses) *	0.026	236.23	255.46
	Demographics			
	PRC ~ BMI	0.002	233.02	239.43
	PRC ~ Age	<0.001	233.31	239.72
	PRC ~ Gender	0.010	231.02	237.43
	PRC ~ Exercise + Smoking + Alcohol	<0.001	236.69	249.51
	Psychological measures			
	PRC ~ Cognitive reflection task (CRT)	<0.001	192.18	198.16
	PRC ~ Trait self-control (TSC)	<0.001	212.33	218.52
	PRC ~ Eysenck Personality Questionnaire (EPQ) – Extraversion (E) + Neuroticism (N) + Psychoticism (P) + Social Desirability (SD)	0.04	177.35	191.92

	PRC ~ BMI + Age + Gender + CRT + TSC + EPQ-E + EPQ-N + EPQ-P + EPQ-SD	0.05	188.10	220.14
Timing	Logistic regression			
	Economic preferences			
	TIMING ~ Loss aversion**	<0.001	255.48	261.89
	TIMING ~ Utility curvature (gains)*	0.017	251.71	258.12
	TIMING ~ Utility curvature (losses)*	0.011	252.80	259.21
	TIMING ~ Present bias (gains)*	0.007	252.92	259.31
	TIMING ~ **Present bias (losses)** *Note: Present Bias (losses) was only significant after controlling for demographics*	0.018	252.17	285.58
	TIMING ~ Loss aversion + Utility curvature (gains) + Utility curvature (losses) + Present bias (gains) + **Present bias (losses)** *Note: Present Bias (losses) was significant after controlling for demographics*	0.059	250.36	269.55
	Demographics			
	TIMING ~ BMI	0.017	252.23	258.64
	TIMING ~ Age	<0.001	255.49	261.90
	TIMING ~ Gender	<0.001	255.48	261.89
	TIMING ~ Exercise + Smoking + Alcohol	0.010	257.80	270.61
	Psychological measures			
	TIMING ~ Cognitive reflection task (CRT)	0.020	205.21	211.19
	TIMING ~ Trait self-control (TSC)	0.010	227.67	233.85
	TIMING ~ Eysenck Personality Questionnaire (EPQ) – Extraversion (E) + Neuroticism (N) + Psychoticism (P) + Social Desirability (SD)	0.020	195.48	210.04
	TIMING ~ BMI + Age + Gender + CRT +TSC+EPQ-E + EPQ-N + EPQ-P + EPQ-SD	0.070	199.64	231.68
Sequence	Linear regression			
	Economic preferences			
	SEQUENCE ~ Loss aversion*	0.001	427.88	437.49
	SEQUENCE ~ Utility curvature (gains)*	0.007	426.87	436.48
	SEQUENCE ~ Utility curvature (losses)*	0.001	427.98	437.59
	SEQUENCE ~ Discounting (gains)*	0.011	426.14	435.75
	SEQUENCE ~ Discounting (losses)*	0.001	427.93	437.54

	SEQUENCE ~ Loss aversion + Utility curvature (gains) + Utility curvature (losses) + Discounting (gains)+ Discounting (losses)*	0.022	432.09	454.52
	Demographics			
	SEQUENCE ~ BMI	0.009	426.51	436.12
	SEQUENCE ~ Age	0.010	426.22	435.83
	SEQUENCE ~ Gender	<0.001	428.09	437.71
	SEQUENCE ~ Exercise + Smoking + Alcohol	0.009	430.48	446.50
	Psychological measures			
	SEQUENCE ~ Cognitive reflection task (CRT)	<0.001	340.59	349.56
	SEQUENCE ~ Cognitive reflection task (TSC)	<0.001	371.62	380.91
	SEQUENCE ~ Eysenck Personality Questionnaire (EPQ) – Extraversion (E) + Neuroticism (N) + Psychoticism (P) + Social Desirability (SD)	0.030	315.68	333.16
	SEQUENCE ~ BMI + **Age +** Gender + CRT +TSC+ EPQ-E + EPQ-N + EPQ-P + EPQ-SD	0.110	315.78	350.73
Risk	Linear regression			
	Economic preferences			
	RISK ~ Loss aversion*	0.001	1685.40	1695.01
	RISK ~ Probability weighting (gains)*	0.008	1684.09	1693.70
	RISK ~ Probability weighting (losses)*	0.011	1683.45	1693.06
	RISK ~ Loss aversion + Utility curvature (losses) + Probability weighting (losses) + Present bias (losses) + Discounting (losses)*	0.020	1685.78	1701.80
	Demographics			
	RISK ~ BMI	0.003	1684.95	1694.56
	RISK ~ Age	0.001	1685.39	1695.00
	RISK ~ Gender	<0.001	1685.44	1695.05
	RISK ~ Exercise + Smoking + Alcohol	0.010	1687.60	1703.62
	Psychological measures			
	RISK ~ Cognitive reflection task (CRT)	0.010	1372.68	1381.65
	RISK ~ Trait self-control (TSC)	<0.001	1517.89	1287.15
	RISK ~ Eysenck Personality Questionnaire (EPQ) – Extraversion (E) + Neuroticism (N) + Psychoticism (P) + Social Desirability (SD)	0.010	1269.67	1287.15
	RISK ~ BMI + Age + Gender + CRT + TSC + EPQ-E + EPQ-N + EPQ-P + EPQ-SD	0.030	1277.66	1312.61

For the Timing dimension, after exploring many different model specifications, one of the better fitting models included: present bias for losses, Eysenck Personality Questionnaire dimensions: Neuroticism and Psychoticism, Cognitive Reflection and BMI. Logistic regression results are reported in Table O.5.C.2.

Table O.5.C.2. Results for exploratory logistic regression for timing dimension.

Predictor	**Estimate**	**SE**	**Z-value**	***p* value**
(Intercept)	-0.46	1.42	-0.33	0.74
Present bias (losses)	1.65	0.94	1.76	0.08
Eysenck Personality Questionnaire (Neuroticism)	1.71	0.97	1.76	0.08
Eysenck Personality Questionnaire (Psychoticism)	0.98	1.02	0.97	0.34
Cognitive reflection	-0.38	0.16	-2.33	0.02
BMI	-0.09	0.04	-1.80	0.07

These results indicate that: those with weaker present bias for losses (marginally significant), those are more prone to neuroticism (marginally significant), those who more on their automatic system, and those with a lower BMI (marginally significant) are more likely to choose a weekly pay-out structure.

Appendix Chapter 6

This Appendix chapter contains a total of two appendices relevant to Chapter 6 and provide information on task instructions, including screenshots of the visualizations of TTO, SG and the validation task (Appendix A), and results after excluding respondents who violate logical consistency (Appendix B).

Abbreviations:

TTO	Time trade-off
SG	Standard gamble
QALY	quality-adjusted life year

Appendix A. Task instructions and screenshots for TTO, SG and validation

Figures O.6.A.1 to O.6.A.3 show screenshots obtained directly from the experimental program (available on request), and portray the instructions used to introduce TTO, SG and the validation task.

Task instructions

Figure O.6.A.1: Screenshot of task instruction for TTO

Instructions

Setting: Imagine that today you are told that you have 10 more years left to live, which you will spend with reduced quality of life. For example, you may experience problems in one or more dimensions, such as mobility, self-care, usual activities, or you may feel pain or anxiety. Throughout this experiment, you will be instructed to imagine living those 10 years in one of several health states (called V, W, X, Y and Z).

Treatment choice task: However, you are also offered a choice to get treatment that improves your quality of life, you can choose to give up some life years to live in full health (i.e. no problems on any dimensions) for a shorter period. In this part of the experiment, your task involves indicating your preference between remaining for 10 years in the described health state (Option A), or living in full health (i.e. a state without any problems) for different durations (Option B). Throughout, you can find reminders of the descriptions of the effects and health states below the choice options.

After you indicate your preference, the duration in full health in Option B will change (sometimes slightly), so please pay careful attention to the treatment outcomes for each choice. After several choices, you will be asked to indicate for which duration in perfect health in Option B, you would be indifferent between both options. In other words, the questions is: from which duration in perfect health would you be as happy (or unhappy) with receiving either Option A or B? If Option A and B are at some point of exactly equal value to you, you can also click a button that to indicate that you are indeed indifferent between Option A and B.

To familiarize you with the procedure, we will start with an example, in which the health state in which you will live for 10 more years represents living with chronic back pain.

Click the tab to start with the practice choice task

Figure O.6.A.2: Screenshot of task instruction for SG

Instructions

Setting: Imagine that today you are told that you have 10 more years left to live, which you will spend with reduced quality of life. For example, you may experience problems in one or more dimensions, such as mobility, self-care, usual activities, or you may feel pain or anxiety. Throughout this experiment, you will be instructed to imagine living those 10 years in one of several health states (called V, W, X, Y and Z).

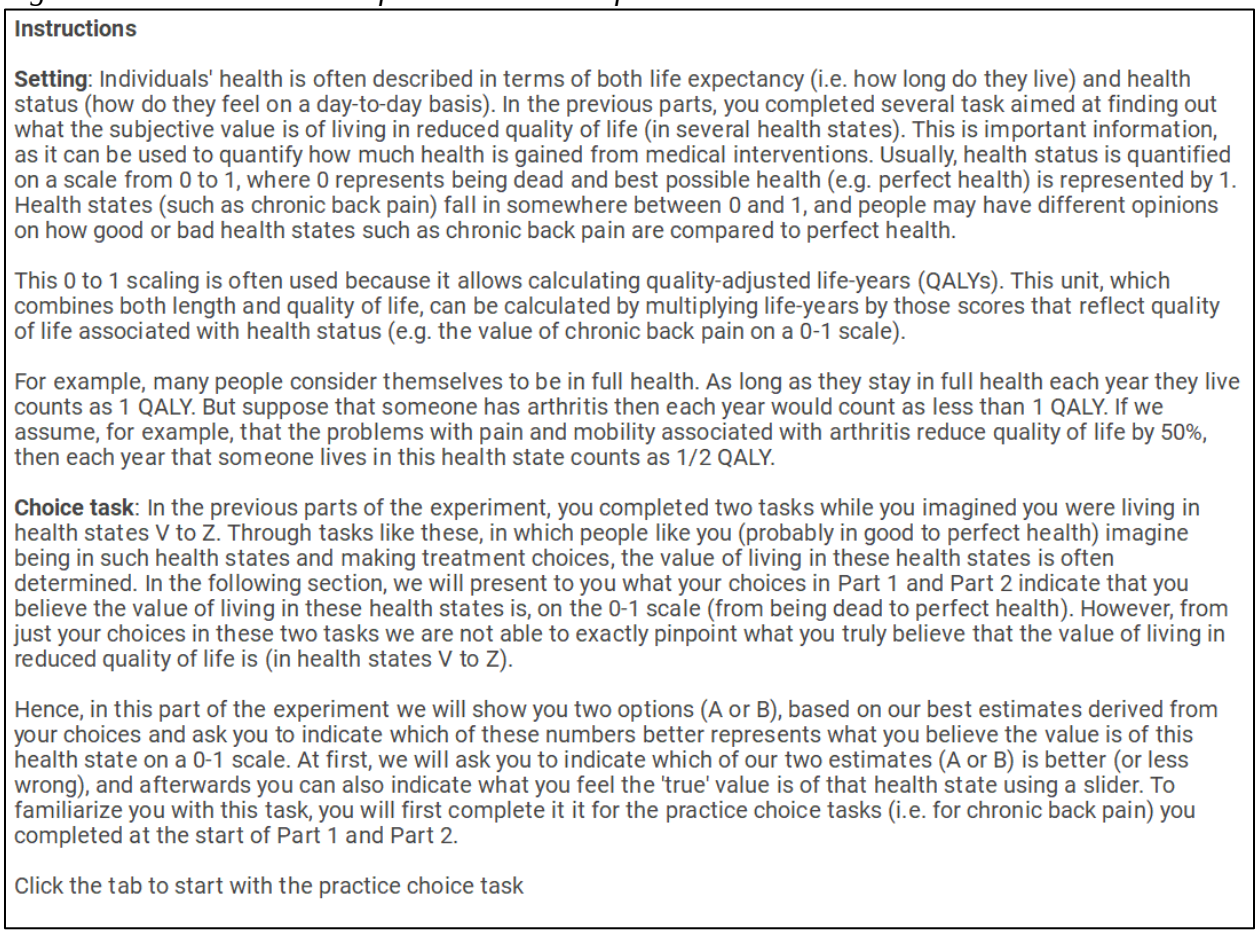
Treatment choice task: However, you are also offered a choice to get treatment that improves your quality of life, you can choose to give up some life years to live in full health (i.e. no problems on any dimensions) for a shorter period. In this part of the experiment, your task involves indicating your preference between remaining for 10 years in the described health state (Option A), or living in full health (i.e. a state without any problems) for different durations (Option B). Throughout, you can find reminders of the descriptions of the effects and health states below the choice options.

After you indicate your preference, the duration in full health in Option B will change (sometimes slightly), so please pay careful attention to the treatment outcomes for each choice. After several choices, you will be asked to indicate for which duration in perfect health in Option B, you would be indifferent between both options. In other words, the questions is: from which duration in perfect health would you be as happy (or unhappy) with receiving either Option A or B? If Option A and B are at some point of exactly equal value to you, you can also click a button that to indicate that you are indeed indifferent between Option A and B.

To familiarize you with the procedure, we will start with an example, in which the health state in which you will live for 10 more years represents living with chronic back pain.

Click the tab to start with the practice choice task

Figure O.6.A.3: Screenshot of task instructions for the validation task



Instructions

Setting: Individuals' health is often described in terms of both life expectancy (i.e. how long do they live) and health status (how do they feel on a day-to-day basis). In the previous parts, you completed several task aimed at finding out what the subjective value is of living in reduced quality of life (in several health states). This is important information, as it can be used to quantify how much health is gained from medical interventions. Usually, health status is quantified on a scale from 0 to 1, where 0 represents being dead and best possible health (e.g. perfect health) is represented by 1. Health states (such as chronic back pain) fall in somewhere between 0 and 1, and people may have different opinions on how good or bad health states such as chronic back pain are compared to perfect health.

This 0 to 1 scaling is often used because it allows calculating quality-adjusted life-years (QALYs). This unit, which combines both length and quality of life, can be calculated by multiplying life-years by those scores that reflect quality of life associated with health status (e.g. the value of chronic back pain on a 0-1 scale).

For example, many people consider themselves to be in full health. As long as they stay in full health each year they live counts as 1 QALY. But suppose that someone has arthritis then each year would count as less than 1 QALY. If we assume, for example, that the problems with pain and mobility associated with arthritis reduce quality of life by 50%, then each year that someone lives in this health state counts as 1/2 QALY.

Choice task: In the previous parts of the experiment, you completed two tasks while you imagined you were living in health states V to Z. Through tasks like these, in which people like you (probably in good to perfect health) imagine being in such health states and making treatment choices, the value of living in these health states is often determined. In the following section, we will present to you what your choices in Part 1 and Part 2 indicate that you believe the value of living in these health states is, on the 0-1 scale (from being dead to perfect health). However, from just your choices in these two tasks we are not able to exactly pinpoint what you truly believe that the value of living in reduced quality of life is (in health states V to Z).

Hence, in this part of the experiment we will show you two options (A or B), based on our best estimates derived from your choices and ask you to indicate which of these numbers better represents what you believe the value is of this health state on a 0-1 scale. At first, we will ask you to indicate which of our two estimates (A or B) is better (or less wrong), and afterwards you can also indicate what you feel the 'true' value is of that health state using a slider. To familiarize you with this task, you will first complete it it for the practice choice tasks (i.e. for chronic back pain) you completed at the start of Part 1 and Part 2.

Click the tab to start with the practice choice task

Figures O.6.A.4 to O.6.A.6 show screenshots obtained directly from the experimental program portraying the visualisation of all three tasks.

Figure O.6.A.4: Screenshot of choice screen for TTO task

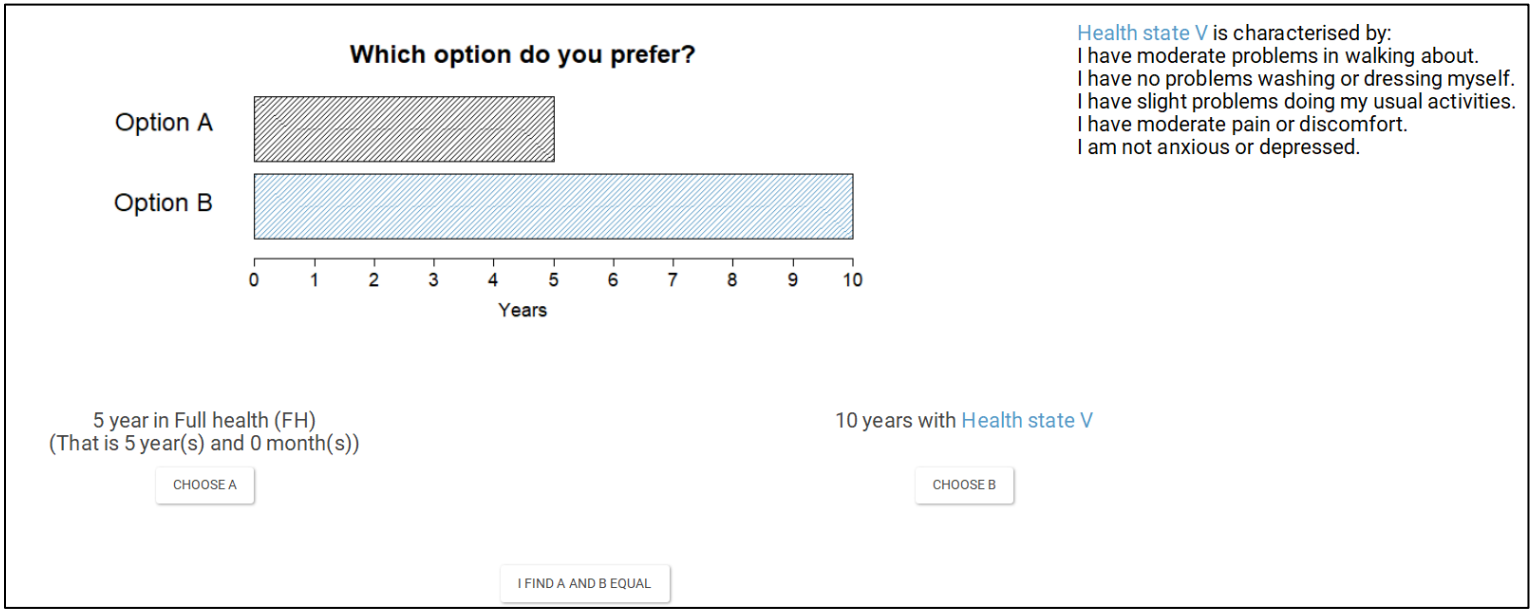


Figure O.6.A.5: Screenshot of choice screen for SG task

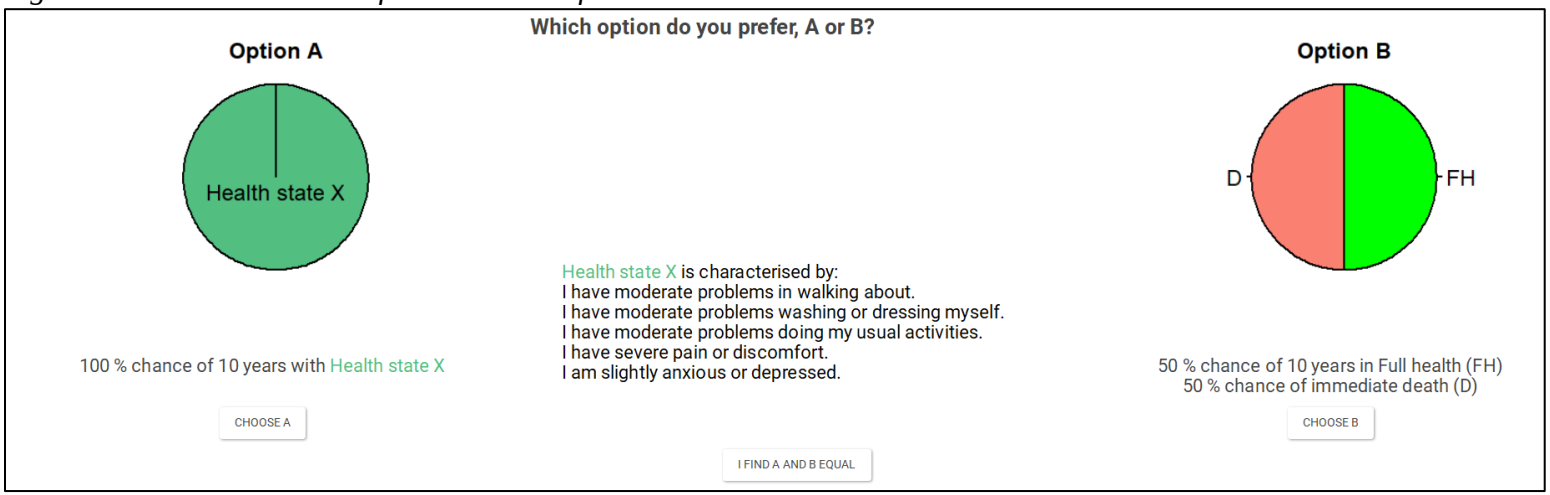
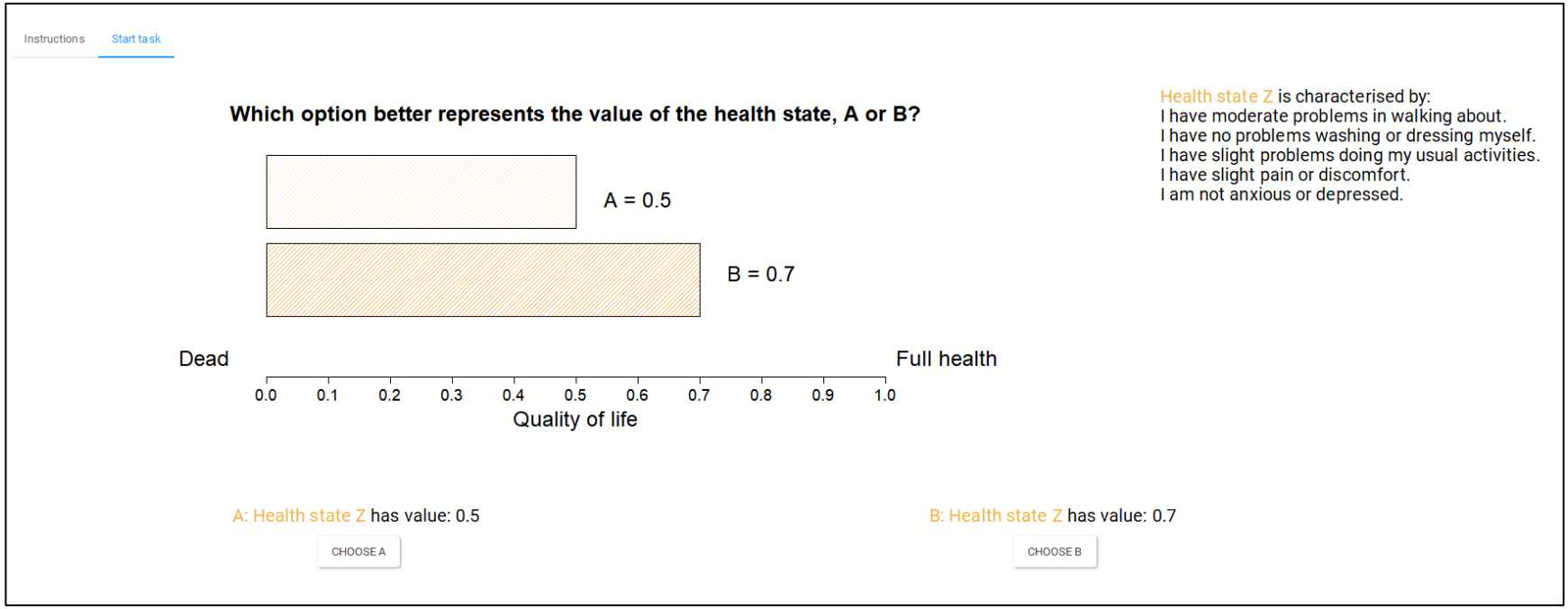


Figure O.6.A.6: Screenshot of choice screen for the validation task



Appendix B. Results after excluding logically inconsistent respondents

The health states used in this experiment were selected to enable a test of logical consistency, as health states were dominating alternatives of each other. This means, that moving from a milder to a more severe health states always yielded more (or at least the same) problems on all dimensions. This implies that if a respondent prefers years with more quality of life to those with less (i.e. monotonicity), health states' valuations should be monotonically decreasing in severity. If this is not the case, this leads to inconsistencies. For example, a respondent reports QALY weights assigned to 31221 (Q2) that are lower than those for 31231 (Q3). This is an inconsistency as it suggests the respondent prefers less health to more.

Throughout this appendix, we will define a respondent as inconsistent when QALY weight Q_i < QALY weight Q_{i+1} (with i = 1 to 4), i.e. when a more severe health state receives a strictly higher QALY weight. Such inconsistencies could occur for each task, i.e. after valuing all health states with TTO, SG and providing a validated QALY weight. Given that only 13% of our sample show such consistency for all health states with both TTO, SG and after validation, we chose not to exclude any respondents in the main text. When comparing consistency between methods, we find significantly fewer inconsistencies after validation compared to both methods (paired Wilcoxon test, $p < 0.03$).

In this Appendix we will reprint Figure 6.1 and Table 6.2 (in Chapter 6 of this dissertation) under several exclusion rules, which differ in terms of how lenient they are. Given that we use relatively mild health states, some noise in elicitation may yield inconsistent QALY weights; hence, in some exclusion rules we allow respondents to be inconsistent for 1 health state.

Exclusion rule 1 (strict): Only those respondents are included that are fully consistent across all three tasks (n = 15).

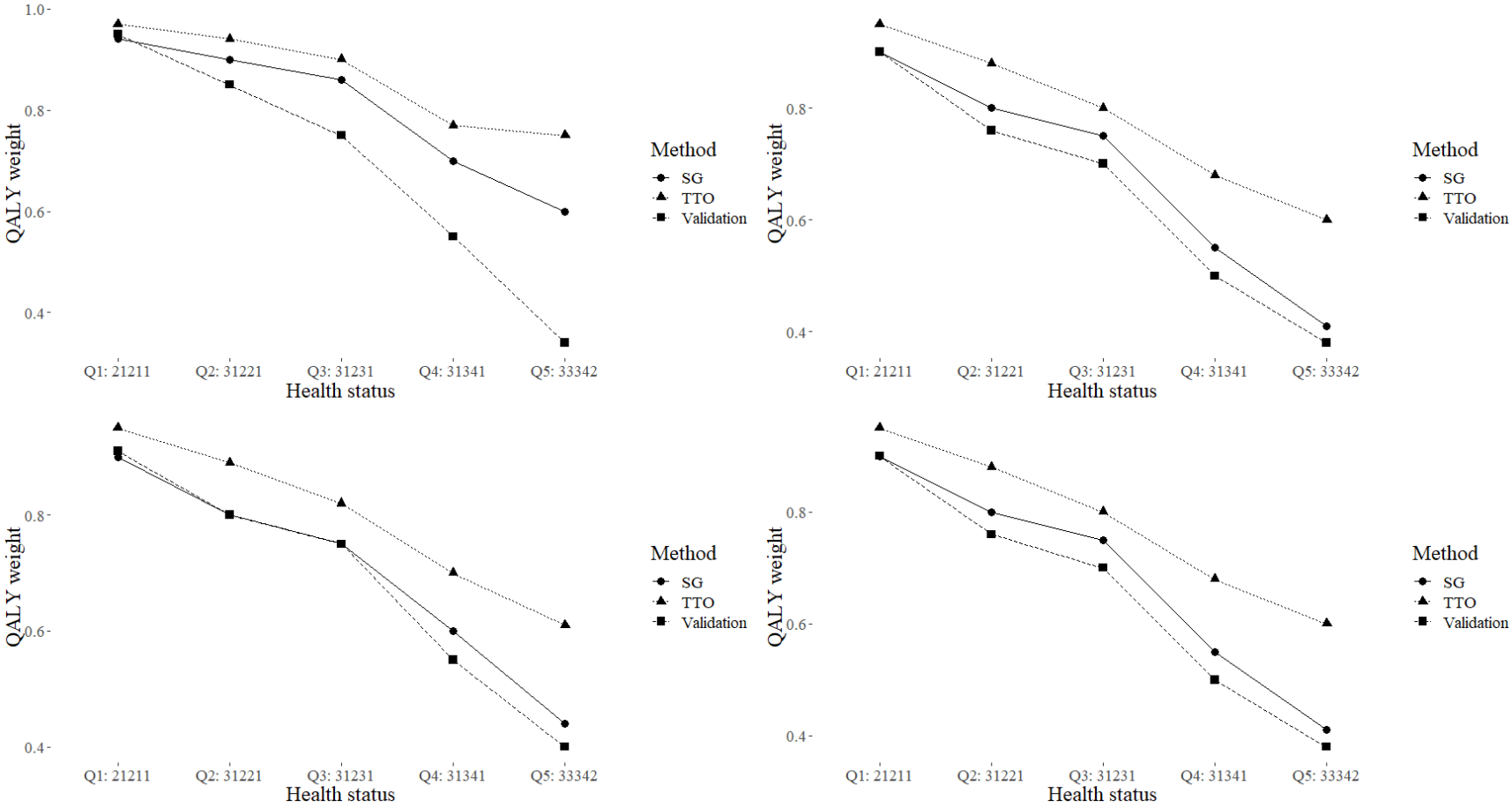
Exclusion rule 2 (strict within-task): For each task, only respondents are included that are consistent in that task, i.e. being inconsistent in one task only leads to exclusion for analysis of that task (TTO: n = 45, SG: n = 102, validation: n = 109).

Exclusion rule 3 (strict with noise): Only those respondents are included that have no more than 1 inconsistency per task for all three tasks (n = 85).

Exclusion rule 4 (strict within-task with noise): For each task separately, only those respondents are included that have no more than 1 inconsistency in that task (TTO: n = 102, SG: n = 102, validation: n = 109).

The remaining samples under each inclusion rule show that especially enforcing a strict exclusion rule (rule 1) leads to a large reduction in sample, which appears mostly a result of low strict consistency for TTO.

Figure O.6.B.1: Figure 6.1 under Exclusion rule 1 (upper left), Exclusion rule 2 (upper right), Exclusion Rule 3 (lower left) and Exclusion rule 4 (lower right)



As can be seen from these reprints, under all exclusion rules the general conclusion remains the same, i.e. QALY weights are higher for SG than for TTO and decrease somewhat when validated.

Table O.6.B.2 shows which methods respondents indicated to better represent the value of each health state under each exclusion rule. Again, under all exclusion rules, the majority indicates that TTO better represents the value of health states (with the exception of Q1 under exclusion rule 3). Overall, these proportions are significant for all samples, and under exclusion rule 2 to 4 we find within health states differences to be significant.

Table O.6.B.2: Number of respondents that indicate to prefer one method over the other in the valuation task.

	Exclusion rule 1		Exclusion rule 2		Exclusion rule 3		Exclusion rule 4	
	TTO	SG	TTO	SG	TTO	SG	TTO	SG
Q1: 21211	9	6	62	56	44	41	62	56
Q2: 31221	10	5	66	52	34	51	66	52
Q3: 31231	9	6	63	55	45	40	63	55
Q4: 31341	10	5	78*	40	60*	25	78*	40
Q5: 33342	10	5	78*	40	64*	21	78*	40
Total	48*	27	347*	243	247*	178	347*	243

Note: * signifies that this percentage is significantly higher with Chi-squared test $p<0.05$,

+ the distribution for Rule 2 and Rule 4 is identical, as allowing noise into the valuation task did not change which respondents were included.

Appendix Chapter 7

This Appendix chapter contains a total of four appendices relevant to Chapter 7, which contain an overview of earlier work on correcting for bias (Appendix A), proofs for the equations used for correcting time trade-off and standard gamble (Appendix B), an overview of the experiment (Appendix C), a formal description of the non-parametric method (Appendix D), a reprinting of task instructions (Appendix E), and additional results for corrections based on parametric assumptions (Appendix F).

Abbreviations:

TTO	Time trade-off
SG	Standard gamble
QALY	quality-adjusted life year
FH	Full health

Appendix A: Overview of literature on correction for TTO and SG

Table O.7.A.1. Overview of studies applying corrections to TTO and/or SG, with differences between methodologies and results categorized.

Authors (by year)	Experiment	Measure corrected	Corrections	(Non-) Parametric	Effect
Stiggelbout et al. (1994)	N=31/D=2	TTO	UC	PAR	TTO: +
Bleichrodt et al. (1999)	N=172/D=1	SG	PW	AS	SG: -
Martin et al. (2000)	N=199/D=3	TTO	UC	PAR	TTO: -
Bleichrodt et al. (2001)	N=51/D=2	SG	LA/PW	AS	SG: -
Bleichrodt (2001)	N=66/D=4	SG	PW	AS	SG -
Oliver (2003)	N=30/D=3	SG	LA/PW	AS	SG: -
Van Osch et al. (2004)	N=45/D=6	both	*SG*: LA/PW	SG: AS	SG: -
			TTO: UC	TTO: PAR	TTO: =
Van der Pol & Roux (2005)	N= 111/D= 1	TTO	UC	AS/PAR	TTO: +
Bleichrodt et al. (2007)	N=65/D=2	SG	LA/PW	AS	SG: -
Attema & Brouwer (2008)	N=70/D=2	TTO	UC	NPAR	TTO: +
Attema & Brouwer (2009)	N=70/D=2	TTO	UC	NPAR	TTO: +
Abellán-Perpiñán et al. (2009)	N=80[†] /D=2	TTO	UC	PAR	TTO: -
Attema & Brouwer (2010)	N=70/D=1	TTO	UC	NPAR	TTO: +
Doctor et al. (2010)	Meta-analysis	SG	LA/PW	AS	SG: -
Attema & Brouwer (2012a)	N=83/D=1	TTO	UC	NPAR	TTO: +
Attema & Brouwer (2012b)	N=83/D=1	TTO	UC	NPAR	TTO: +
Attema et al. (2013b)	N=159/D=6	TTO	UC	NPAR	TTO: =
Pinto-Prades & Abellán-Perpiñán, (2012)	N=65/D=2	SG	LA/PW	AS[‡] [§]	SG: -
Attema & Brouwer (2014)	N=80-112 /D=5	TTO	UC	NPAR	TTO: +

Note: *N* = sample size, *D* = amount of health states, and the following abbreviations are used: LA (loss aversion), PW (probability weighting), UC (utility curvature or time preference), AS (assumptions about parameter estimates), PAR (parametric elicitation), NPAR (non-parametric elicitation), + (upward effect of correction, and – (downward effect of correction). [†] = between-subjects tariff estimation with total N=720 and total D=18, [‡] = parameters were obtained through median optimization. [§]= based on the same data as Bleichrodt et al. (2007).

Appendix B: Proofs for correction of TTO and SG

Below we will provide proofs for Eq. 7.3 and Eq. 7.4 (in Chapter 7 of the dissertation). We will first show how TTO and SG indifferences are derived under the general QALY model, and show how without loss of generality we can rewrite these derivations under the general QALY model. Next we apply our extension of this model as described in Chapter 7, to obtain proofs for Eq. 7.3 and Eq. 7.4. Comparing proofs for rewritten derivations under the general QALY model and the derivation of Eq. 7.3 and Eq. 7.4 show how the general QALY model is a special case of our extended model based on PT is.

TTO and SG weights in the general QALY model

TTO typically involves indifferences of the form: $(\beta_x, T_x) \sim \left(FH, T_y\right)$. If we apply the general QALY model this indifference is evaluated by:

$$U(\beta_x)L(T_x) = U(FH)L(T_y), \tag{O7B1}$$

with $U(FH) = 1$, this allows derivation of TTO weights (i.e. $U(\beta_x)$), by:

$$U(\beta_x) = L(T_y)/L(T_x). \tag{O7B2}$$

SG typically involves indifferences of the form: $(\beta_x, T_x) \sim (FH, T_x)_p(D)$. If we apply the general QALY model (assuming EU), this indifference is evaluated by:

$$U(\beta_x)L(T_x) = pU(FH)L(T_x) + (1-p)D, \tag{O7B3}$$

with $U(FH) = 1$ and $U(D) = 0$. This allows derivation of SG weights (i.e. $U(\beta_x)$), by:

$$U(\beta_x) = p. \tag{O7B4}$$

Proof for Equation 7.3 – TTO

The first extension of the general QALY model we apply is reference-dependence, which is the most fundamental break-away from the general QALY model. To derive Eq. 7.3, we assume that the time spent in reduced health status is the RP, i.e. $(\beta_r, T_r) = (\beta_x, T_x)$ in scenario A. Compared to this RP, the time spent in full health yields improved health status but a loss in life duration. Hence, the indifference in scenario A, $(\beta_x, T_x) \sim \left(FH, T_y\right)$, implies indifference between staying at the RP and an improvement in health status experienced for (T_y) years and a loss in life duration $\left(T_y - T_x\right)$. Importantly, duration is determined relative to the RP (T_r). Hence, after incorporating reference-dependence the indifference $(\beta_x, T_x) \sim \left(FH, T_y\right)$ is evaluated by:

$$U(\beta_x)L(T_x - T_r) = \left(U(FH) - U(\beta_x)\right)\left(L\left(T_y - T_r\right) - L(T_a - T_r)\right) + U(\beta_x)\left(L\left(T_y - T_r\right) - L(T_x - T_r)\right). \tag{O7B5}$$

Next, we apply our notational convenience of writing $T_x^* = T_x - T_r$, for any T. Furthermore, we incorporate sign-dependence (i.e. $L^i(T)$) as described in Section 2, Because compared to T_x^*, durations T_y^* and T_a^* are losses, this yields:

$$U(\beta_x)L^+(T_x^*) = \left(U(FH) - U(\beta_x)\right)\left(L^-\left(T_y^*\right) - L^-(T_a^*)\right) + U(\beta_x)\left(L^-\left(T_y^*\right) - L^+(T_x^*)\right), \tag{O7B6}$$

Next, we multiply the loss in life duration $\left(T_y - T_x\right)$ with the loss aversion index to obtain:

$$U(\beta_x)L^+(T_x^*) = \left(U(FH) - U(\beta_x)\right)\left(L^-\left(T_y^*\right) - L^-(T_a^*)\right) + U(\beta_x)\lambda\left(L^-\left(T_y^*\right) - L^+(T_x^*)\right). \quad \text{(O7B7)}$$

We normalize throughout, as is common in applications of PT, such that the utility of staying at the RP is zero, i.e. $L^+(T_x^*) = 0$, and the utility of full health is one, $U(FH) = 1$, which yields:

$$\left(1 - U(\beta_x)\right)\left(L^-\left(T_y^*\right) - L^-(T_a^*)\right) = -\lambda U(\beta_x)L^-\left(T_y^*\right). \quad \text{(O7B8)}$$

Rearranging Eq. O7B8 gives:

$$\frac{1-U(\beta_x)}{U(\beta_x)} = \frac{-\lambda L^-\left(T_y^*\right)}{L^-\left(T_y^*\right) - L^-(T_a^*)}. \quad \text{(O7B9)}$$

Solving for $U(\beta_x)$ gives:

$$U(\beta_x) = \frac{L^-\left(T_y^*\right) - L^-(T_a^*)}{(1-\lambda)L^-\left(T_y^*\right) - L^-(T_a^*)}, \quad \text{(O7B10)}$$

which gives Eq. 7.3 if we set $L^-(T_a^*) = -1$.

Importantly, under the general QALY model we can derive the same derivation for $U(\beta_x)$ as follows: we take Eq. O7B1, subtract $U(\beta_x)L(T_x)$ from both sides to obtain $0 = U(FH)L(T_y)$- $U(\beta_x)L(T_x)$. Adding $U(\beta_x)L\left(T_y\right)$ to both sides and rearranging yields $0 = \left(U(FH) - U(\beta_x)\right)L(T_y)$+ $U(\beta_x)\left(L\left(T_y\right) - L(T_x)\right)$. Realizing that $T_x^* = T_x - T_r$, that $L\left(T_y\right) = L\left(T_y^*\right) - L(T_a^*)$ and that $L(T_x) = L(T_x^*) - L(T_a^*)$, gives Eq. O7B8 with $\lambda = 1$ and $L(T) = L^+(T) = L^-(T)$. Simple rearranging will give Eq. O7B10 , which shows that the general QALY model is a special case of the reference-dependent evaluation.

Proof for Equation 7.4 - SG

We assume (β_x, T_x) is the reference point (i.e. $(\beta_r, T_r) = (\beta_x, T_x)$). If we introduce reference-dependence, the gamble $(FH, T_x)_p(D)$ then consists of the option to gain quality of life for $T_x - T_a$ years with a probability p, and a risk of death after T_a years with probability $1 - p$, which is a loss in lifetime of $T_a - T_x$ years in health state β_x. Again, these durations are determined relative to T_r.Thus, when the respondent is indifferent between the two options, the following equation holds:

$$U(\beta_x)L(T_x - T_r) = p(U(FH) - U(\beta_x))\left(L(T_x - T_r) - L(T_a - T_r)\right) + (1-p)U(\beta_x)(L(T_a - T_r) - L(T_x - T_r)). \quad \text{(O7B11)}$$

Second, we introduce probability weighting, by assigning sign-dependent decision weights (i.e. $w^i(p)$) to probabilities, i.e.:

$$U(\beta_x)L(T_x - T_r) = w^+(p)(U(FH) - U(\beta_x))\left(L(T_x - T_r) - L(T_a - T_r)\right) + w^-(1-p)U(\beta_x)(L(T_a - T_r) - L(T_x - T_r)).$$

Next, we incorporate sign-dependence for $L^i(T)$ as described in Chapter 7, and apply our notational convenience of writing $T_x^* = T_x - T_r$, for any T. Because compared to T_x^*, duration T_a^* is a loss, this yields:

$$U(\beta_x)L^+(T_x^*) = w^+(p)(U(FH) - U(\beta_x))\big(L^+(T_x^*) - L^-(T_a^*)\big) + w^-(1-p)U(\beta_x)\big(L^-(T_a^*) - L^+(T_x^*)\big). \quad \text{(O7B12)}$$

Next, we multiply the loss in life duration (i.e. immediate death) with loss aversion to obtain:

$$U(\beta_x)L^+(T_x^*) = w^+(p)\big(U(FH) - U(\beta_x)\big)\big(L^+(T_x^*) - L^-(T_a^*)\big) + w^-(1-p)\,U(\beta_x)\lambda(L^-(T_a^*) - L^+(T_x^*)). \quad \text{(O7B13)}$$

We assume throughout, as is common in applications of PT, that the utility of staying at the RP, i.e. $L^+(T_x^*) = 0$, and $U(FH) = 1$ which yields, after rearranging:

$$\big(1 - U(\beta_x)\big)w^+(p)\big(-L^-(T_a^*)\big) = -\lambda U(\beta_x)w^-(1-p)(L^-(T_a^*)). \quad \text{(O7B14)}$$

Rearranging Eq. O7B14 gives:

$$\frac{1-U(\beta_x)}{U(\beta_x)} = \frac{-\lambda w^-(1-p)L^-(T_a^*)}{w^+(p)\big(-L^-(T_a^*)\big)}. \quad \text{(O7B15)}$$

Solving for $U(\beta_x)$ gives:

$$U(\beta_x) = \frac{w^+(p)\big(-L^-(T_a^*)\big)}{-\lambda w^-(1-p)L^-(T_a^*) + w^+(p)\big(-L^-(T_a^*)\big)}, \quad \text{(O7B16)}$$

which gives Eq. 7.4 if we set $L^-(T_a^*) = -1$.

Again, we can obtain a similar derivation for $U(\beta_x)$ under the general QALY: we take Eq. O7B3A, subtract $U(\beta_x)L(T_x)$ from both sides to obtain $0 = pU(FH)L(T_x)$ - $U(\beta_x)L(T_x)$. Adding $pU(\beta_x)L(T_x)$ to both sides and rearranging yields $0 = p\big(1 - U(\beta_x)\big)L(T_x) - (1-p)U(\beta_x)\big(L(T_x)\big)$. Realizing that $T_x = -T_a^*$, gives Eq. O7B14 with $\lambda = 1$ and $w^+(p) = w^-(p) = p$. Rearranging then gives Eq. O7B16, which shows how the general QALY model is a special case of the reference-dependent evaluation.

On the scaling of $L^-(T_a^*)$

We have shown how Eq. 7.3 and Eq. 7.4 can be derived from Eq. O7B8 and O7B16 respectively, scaling such that $L^-(T_a^*) = -1$. Given that $L^i(T^*)$ is cardinal, we are allowed to freely set the utility of any two points along this scale. We already, as in conventional in applications of PT, set the utility of the RP to 0, i.e. $L^+(T_x^*) = 0$, and, hence, we are allowed to specify the utility of 1 other point. Given that T_a is the lowest possible outcome, we believe it is natural to assign this outcome $L^-(T_a^*) = -1$. This scaling yields Eq. 7.3 and Eq. 7.4 but it is straightforward to show that this scaling is immaterial to the derivation of TTO and SG weights (it matters only for utility elicitation). Take, for example, a subject with $\lambda = 2$, $p = 0.8$, $w^+(0.8) = 0.9$, $w^-(1-0.8) = 0.25$, and $L^-\big(T_y^*\big) = -0.5$. Multiplication of $L^i(T^*)$ by any non-zero constant does not affect TTO or SG weights, and any addition or subtraction satisfying $L^-(T_a^*) \neq 0$, leaves TTO or SG weights unaffected.

Appendix C: Overview of experiment and counterbalancing procedures

First, the order of the two parts of the experiment (*health state valuation vs. non-parametric method*) was counterbalanced. Within the utility elicitation part, the order in which participants faced gains or losses was randomized; half of the participants always completed gain sections first, whilst the other half completed loss sections first. Furthermore, within the health state valuation part, the order in which TTO and SG were presented was randomized, as was the order of the health states. Furthermore, a total of five practice blocks and four consistency checks were used.

Part 1: Health state valuation	**Part 2: Non-parametric method**
Practice TTO (β_p)	*Practice block 1:* Certainty equivalents
TTO (β_1, β_2, β_3)	*Stage 1*: *connecting* $L^+(T^*)$ *and* $L^-(T^*)$
Practice SG (β_p)	*Practice block 2:* Trade-off method
SG (β_1, β_2, β_3)	*Stage 2*: *elicitation of* $L^+(T^*)$
	Consistency block: Re-elicitation of x_2^+
	Stage 3: elicitation of $L^-(T^*)$
	Consistency block: Re-elicitation of x_2^-
	Stage 4a: probability weighting gains $w^+(p)$
	Consistency block: Re-elicitation of randomly selected x_p^+
	Stage 4b: probability weighting losses $w^+(p)$
	Consistency block: Re-elicitation of randomly selected x_p^+

Appendix D: Elaborate formal description of measurement method

Abdellaoui and colleagues (2016) describe the following three-stage methodological procedure, here appended to reflect decision under risk (see Table O7D1 for stimuli used in this adaptation of their method). Unless otherwise specified, the assumptions presented in Chapter 7 will hold. Importantly, we will simplify notation in this Appendix, to ensure comparability to the original authors' work (Abdellaoui et al., 2016)**.** All outcomes refer to health profiles of the form (β, T), which are evaluated as in Eq. 7.1 and Eq. 7.2. Prospects are rank-ordered for duration, which indicates that for gain outcomes $(\beta_x, T_x^*)_p(\beta_y, T_y^*)$ signifies that $T_x \geq T_y$ and for loss outcomes it signifies that $T_x^* < T_y^*$. For prospects containing gain and loss outcomes (i.e. mixed prospects), $(\beta_x, T_x^*)_p(\beta_y, T_y^*)$ signifies that T_x^* is a gain and T_y^* is a loss. Throughout the whole application of this method, β_x will remain constant (i.e. at full health), and will thus cancel out in the evaluation of each indifference mentioned in this Appendix. For example the indifference, $(\beta_x, T_x^*)_p\left(\beta_x, T_y^*\right) \sim (\beta_x, T_z^*)$, with $T_y^* < 0 < T_z^* < T_x^*$, will be evaluated as:

$$w^+(p)U(\beta_x)L^+(T_x^*) + w^-(1-p)U(\beta_x)L^-\left(T_y^*\right) = U(\beta_x)L^+(T_z^*),$$

where dividing by $U(\beta_x)$ gives:

$$w^+(p)L^+(T_x^*) + w^-(1-p)L^-\left(T_y^*\right) = L^+(T_z^*).$$

For brevity, we will suppress $U(\beta_x)$ in the derivations of the non-parametric method from here onwards, and instead apply the general notation also used by Abdellaoui and colleagues (2016). Hence, all outcomes are health profiles, i.e.**:** x_0(reference point), $l, \ell, \mathcal{L}, g, \mathcal{g}, \mathcal{G}$, $[x_1^+, x_2^+, \dots, x_{k_G}^+]$ and $[x_1^-, x_2^-, \dots, x_{k_L}^-]$ refer to T^* life years in β_x.

First stage: connecting $L^+(T^*)$ and $L^-(T^*)$

First, we select a probability *p* that is kept constant throughout the first three stages and a gain g. Next, l is elicited, by means of the following indifference $g_p l \sim x_0$.

Equation (1) thus implies that:

$$w^+(p)L^+(g) + w^-(1-p)L^-(l) = L^+(x_0) = 0. \qquad \text{(O7D1)}$$

Then, the certainty equivalents x_1^+ and x_1^- are elicited, by the following indifferences: $x_1^+ \sim g_p x_0$ and $x_1^- \sim l_p x_0$. These indifferences consequently imply:

$$L^+(x_1^+) = w^+(p)L^+(g), \qquad \text{(O7D2)}$$

and equivalently for losses:

$$L^-(x_1^-) = w^-(p)L^-(l). \qquad \text{(O7D3)}$$

Combining Eqs. (O7D1) – (O7D3) gives:

$$L^+(x_1^+) = -L^-(x_1^-). \qquad \text{(O7D4)}$$

Through this equation we obtain the first elements of the standard sequence (x_1^+ and x_1^-), which is elicited in subsequent stages.

Second and third stage: elicitation of $L^+(T^*)$ and $L^-(T^*)$.

Next, the trade-off method by Wakker and Deneffe (1996) is employed to elicit a standard sequence. Let ℓ be a prespecified loss. First, subjects are presented with the prospects ${x_1^+}_p\mathcal{L}$ and $\ell_p x_0$, in order to elicit the loss $\mathcal{L}$ for which subjects are indifferent (x_1^+ is the gain from stage 1). The indifference ${x_1^+}_p\mathcal{L} \sim \ell_p x_0$ gives:

$$w^+(p)L^+(x_1^+) + w^-(p)L^-(\mathcal{L}) = w^-(p)L^-(\ell). \qquad \text{(O7D5)}$$

Through rearranging Eq. (D5) we obtain,

$$L^+(x_1^+) - L^+(x_0) = \frac{w^-(p)}{w^+(p)}\big(L^-(\ell) - L^-(\mathcal{L})\big). \qquad \text{(O7D6)}$$

Second, subjects are presented with the prospects ${x_2^+}_p\mathcal{L}$ and ${x_1^+}_p\ell$, where the gain x_2^+ is varied such that they are indifferent ${x_2^+}_p\mathcal{L} \sim {x_1^+}_p\ell$. This indifference implies, after rearranging:

$$L^+(x_2^+) - L^+(x_1^+) = \frac{w^-(p)}{w^+(p)}\big(L^-(\ell) - L^-(\mathcal{L})\big). \qquad \text{(O7D7)}$$

Combining Eqs. (O7D6) and (O7D7) gives:

$$L^+(x_2^+) - L^+(x_1^+) = L^+(x_1^+) - L^+(x_0). \qquad \text{(O7D8)}$$

Finally, we elicit a series of indifferences: ${x_j^+}_p\mathcal{L} \sim {x_{j-1}^+}_p\ell, j = 2, \dots, k_G$, which together form the standard sequence $[x_0, x_1^+, x_2^+, \dots, x_{k_G}^+]$ for gains. It follows straightforwardly that for all j, $L^+\big(x_j^+\big) - L^+(x_{j-1}^+) = L^+(x_1^+) - L^+(x_0)$.

The standard sequence for losses is constructed equivalently (part three). Subjects face similar prospects, where we first fix a gain $\mathcal{g}$ to elicit the gain $\mathcal{G}$ such that subjects produce the following indifference: $\mathcal{G}_p x_1^- \sim \mathcal{g}_p x_0$. Finally, we elicit the standard sequence $[x_0, x_1^-, x_2^-, \dots, x_{k_L}^-]$ as described above using the following general form: $\mathcal{G}_p x_j^- \sim \mathcal{g}_p x_{j-1}^-, j = 2, \dots, k_L$.

Fourth stage: probability weights

To measure the probability weighting functions $w^+(p)$ and $w^-(p)$, we asked for the certainty equivalents x_p^+ and x_p^- of the prospects ${x_{kG}^+}_p x_0$ and ${x_{kL}^-}_p x_0$. The outcomes $x_{k_G}^+$ and $x_{k_G}^-$ are the maximum (minimum) outcome elicited in the standard sequence. Therefore, it follows from the probability weighting function and the chosen scaling of utility that $L^+(x_P^+) = w^+(p)$ and $-L^-\big(x_p^-\big) = w^-(p)$. The values of $L^+\big(x_p^+\big)$ and $L^-\big(x_p^-\big)$ are interpolated from

their respective standard sequences (elicited in stage 2 and 3). The probability p was varied (0.1; 0.3; 0.5; 0.7; 0.9) to measure the probability weighting functions for a wide range of probabilities, both for gains and losses.

Table O7D1: Four-stage procedure to measure utility of life duration

The third column shows the variables assessed in each stage, and column four shows the elicited indifferences. The fifth column shows the implication of these elicited indifferences. The final column shows the stimuli used in this experiment.

	Elicited	Indifference	Implication	Stimuli
Stage 1	l	$g_p l \sim x_0$	$L^+(x_1^+) = -L^-(x_1^-)$	$g = 5\ years$ $p = 0.5$ $x_0 = 70\ years$
	x_1^+	$x_1^+ \sim g_p x_0$		
	x_1^-	$x_1^- \sim l_p x_0$		
Stage 2	$\mathcal{L}$	${x_1^+}_p \mathcal{L} \sim \ell_p x_0$	$L^+(x_j^+) - L^+(x_{j-1}^+) = L^+(x_1^+) - L^+(x_0)$	$\ell = -1$ year
	x_j^+	${x_j^+}_p \mathcal{L} \sim {x_{j-1}^+}_p \ell$		
Stage 3	$\mathcal{G}$	$\mathcal{G}_p x_1^- \sim g_p x_0$	$L^-(x_j^-) - L^-(x_{j-1}^-) = L^-(x_1^-) - L^+(x_0)$	$g = 1\ year$
	x_j^-	$\mathcal{G}_p x_j^- \sim g_p x_{j-1}^-$		
Stage 4	x_p^+	$x_p^+ \sim x_{kG_p}^+ x_0$	$L^+(x_p^+)/L^+(x_{kG}^+) = w^+(p)$	$p = \begin{cases} 0.1 \\ 0.3 \\ 0.5 \\ 0.7 \\ 0.9 \end{cases}$
	x_p^-	$x_P^- \sim x_{kL_p}^- x_0$	$L^-(x_p^-)/L^-(x_{kL}^-) = w^-(p)$	$p = \begin{cases} 0.1 \\ 0.3 \\ 0.5 \\ 0.7 \\ 0.9 \end{cases}$

Appendix E: Experimental instructions translated from Dutch and example screenshots.

Utility elicitation

The following instruction was used to introduce the non-parametric method:

'Imagine you will live until 70 years old in perfect health. After becoming 70 you will contract a deadly disease, which will lead to a direct, painless death.

As a result of recent developments in pharmacological sciences, several drugs have become available. Your task involves comparing these drugs, and indicating your preferred drug. In any case you will be comparing two drugs. The drugs are described as Drug A and Drug B. In some cases you will have the opportunity to take a drug that has no effect on your health, as if you would be taking a placebo pill.

For some drugs it is not yet entirely clear if they work, picking these drugs involves some risk. For example, Drug B has 50% chance of increasing your length of life by 4 years, with no adverse effects or other consequences if the Drug fails. It cannot be determined beforehand for whom the drugs will work, you can expect this is determined solely by chance. Personal characteristics have no effect on this chance, so your chances are just as good as any others'.

For some drugs, it is known that if they are ineffective, they could have a negative influence on your health. Using these types of drugs could potentially lead to a reduction in length of life. The result of taking these drugs will be determined immediately by a physician, but its' results are final. No new or better drugs will be developed to change your situation.

After you compared Drug A and B with each other, you will be asked to indicate by means of a slider when you would be indifferent between Drug A and B.'

Stages of non-parametric method

Considering the non-parametric method is chained, i.e. answers from the previous stage carry over to the next meaning that differences could exist between subjects. In Table O.7.E.1, example instructions can be found with some screenshots presenting the visual representation for each stage, where outcomes were likely to be different for each individual subject. Appendix D lists how these outcomes were obtained from revealed preferences.

Table O.7.E.1: Example instructions per stage of the non-parametric method, with indifference elicited (for implications, see Appendix D)

Stage 1: connecting $L^+(T^*)$ and $L^-(T^*)$ – Mixed prospect - $g_p l \sim x_0$
'You have the choice between a remedy (Drug A) that can both increase and decrease your length of life, and a remedy (Drug B) that does not change your situation. If you pick Drug A, you have a chance of 50% to gain 5 life years. If the drug is ineffective, you lose some life years. If you pick Drug B, your situation remains unchanged and you will live until 70.
Stage 1: connecting $L^+(T^*)$ and $L^-(T^*)$ – CE gains - $x_1^+ \sim g_p x_0$

‘Now you have to make a choice between two drugs that will not affect your quality of life, but only your length of life. Drug B will increase your length of life with a couple of years, while if you pick Drug A a 50% chance exists that your length of life will not be increased, with a 50% chance of increasing your length of life with 5 years.
Stage 1: connecting $L^+(T^*)$ and $L^-(T^*)$ – CE losses - $x_1^- \sim l_p x_0$
Again, imagine you would live until 70 in perfect health, after which you would die immediately and painlessly, as described before. However, you have contracted another fatal disease that should also be treated. You have to choose between drugs that will not affect your quality of life, but will affect your length of life to some extent. Two drugs exist, Drug A will reduce your length of life with some years, while if you pick Drug B, a 50% chance exists of reducing your lifetime by **3 years** and a 50% chance that your lifetime will not be reduced at all.
Stage 2: Standard sequence gains $L^+(T^*)$ – Eliciting offset loss - $x_{1\ p}^{+}\mathcal{L} \sim \ell_p x_0$
In this next phase, you compare (as in the practice block) two drugs that both carry a degree of risk. Drug A has a 50% chance of retaining your expected health (70 years), and a 50% chance of you losing 1 year. Drug B has a chance of improving your life expectancy by some years, but may also reduce your lifetime by a few years.
Stage 2: Standard sequence gains $L^+(T^*)$ – Standard sequence - $x_{j\ p}^{+}\mathcal{L} \sim x_{j-1\ p}^{+}\ell$
Now, you will compare several series of drugs with each other. Both drugs may each time both increase your health to a different extent. Additionally, both drugs will have risks of adverse effects, and thus decrease your lifetime. In each series you are asked to pick your preferred drug and specify a value for which you would be indifferent.
Stage 3: Standard sequence losses $L^-(T^*)$ – Eliciting offset gains - $\mathcal{G}_p x_1^- \sim \mathcal{g}_p x_0$
Imagine again that you would become 70 in perfect health, followed by immediate and painless death as a result form the disease described earlier. Yet, you have contracted another disease that is also to be treated. You have to choose between Drugs that do not affect your quality of life, but may affect your length of life. In this phase you compare two drugs for this second disease, which both carry some degree of risk. Drug A has a 50% chance of you retaining your current health (70 years), and a 50% chance of gaining 1 year. Drug B has the chance of improving your length of life, but may also reduce your life time by a few years.
Stage 3: Standard sequence losses $L^-(T^*)$ – Standard sequence - $\mathcal{G}_p x_j^- \sim \mathcal{g}_p x_{j-1}^-$
Next, you will compare series of Drugs with each other. Both drugs may each time both decrease or increase your health. You will be asked which Drug you pick, after which you will be asked to select that value which would make you indifferent between both Drugs.

Stage 4: Probability weighting for gains $w^+(p)$ - $x_p^+ \sim x_{kG_p}^+ x_0$
In this next part, you are again asked to compare different Drugs. Now, however, the drugs may have different chances of success. Your choice is between Drug A with a risky outcome, and Drug B which has a certain outcome. Eventually you will be asked to indicate when you would be indifferent between these Drugs. Please imagine that if your situation does not change, you would live until 70 years old, followed by immediate, painless death.
Stage 4: Probability weighting for gains $w^-(p)$ - $x_p^- \sim x_{kL_p}^- x_0$
Imagine: You would live until 70 years old in perfect health, followed by immediate painless death, as described before. However, you've contracted another unrelated disease that needs to be treated. Your choice in between two Drugs that treat this second disease, which will not affect quality of life, but only your length of life to some extent. One of these has the chance of returning you to your initial expected health (age 70), with a varying chance of success. The other Drug will reduce your life duration with some years for certain.

Health state valuation

The following instruction was used to introduce time trade-off (TTO) and standard gamble (SG):

'In this next part you will compare several Treatments with each other, that will to some extent affect your quality of life. Throughout this part of the experiment, please imagine the following: You will live until age 50 in perfect health. After becoming 50, you will contract a disease, forcing you to choose between two different Treatments, which we will present to you.

You are tasked with comparing treatments and indicating your preferred Treatment. We will describe them as Treatment A and Treatment B. In some cases it is not yet known beforehand if treatments will be effective, and for whom. It cannot be determined beforehand for whom the Treatments will work, you can expect this is determined solely by chance. Personal characteristics have no effect on this chance, so your chances are just as good as any others'. The result of these Treatments will be determined immediately by a physician, but its' results are final. No new or better Treatments will be developed to change your situation.

TTO In this part you compare two Treatments affecting your length of life and quality of life. If you pick Treatment Am you will live for some more years after age 50 in perfect health, followed by immediate, painless death. If you choose Treatment B, you will live 20 more years after age 50 in a reduced quality of life.

SG: In this part you compare two Treatments of which 1 has a certain outcome, while the other has a risky outcome. If you pick the certain, you will live 20 more years after age 50 in a reduced quality of life. The risky treatment offers you the chance to live 20 more years after age 50 in perfect health. However, if this treatment is ineffective, you will die immediately and painlessly after age 50.

Appendix F: Isolated corrections with parametric assumptions

We also performed isolated corrections with parametric assumptions, i.e. where L and w were estimated by means of a power function and Tversky and Kahneman's (Tversky and Kahneman, 1992) one-parameter weighting function respectively. First, we corrected TTO for utility curvature only, with λ = 1. Second, TTO weights were corrected for loss aversion only, with linear utility (i.e. $L^i(T^*) = T^*$). Third, we corrected SG for probability weighting only, with λ = 1. Finally, SG weights were corrected for loss aversion only, with $w^i(p) = p$. This allow us to demonstrate the influence of each correction in isolation. Table S1 shows similar results compared to our findings for non-parametric corrections. For TTO, correcting for loss aversion had a stronger downward influence than correcting for curvature of $L^i(T^*)$. For SG, this finding also holds, i.e. correction for loss aversion had a stronger affect than correcting for probability weighting. Again, this confirms our conclusion that relative ineffectiveness of correction based on parametric estimation appears to be driven by the estimation of $w^i(p)$.

Table O.7.F.1. Isolated effects of corrections for utility curvature (UC), loss aversion (LA) and probability weighting (PW) for TTO and SG weights [standard deviation in brackets].

Health state	Uncorrected weight		UC only		LA only		PW only	
TTO: Implication	$\lambda = 1$ & $L^i(T^*) = T^*$		$\lambda = 1$		$L^i(T^*) = T^*$			
β_1 : 21211	0.665	[0.268]	0.611	[0.296]	0.537	[0.311]		
β_2 : 31221	0.605	[0.259]	0.558	[0.287]	0.474	[0.3]		
β_3 : 32341	0.39	[0.259]	0.365	[0.277]	0.288	[0.259]		
SG: Implication	$\lambda = 1$ & $w^i(p) = p$				$w^i(p) = p$		$\lambda = 1$	
β_1 : 21211	0.75	[0.25]			0.63	[0.307]	0.715	[0.271]
β_2 : 31221	0.706	[0.261]			0.584	[0.305]	0.676	[0.287]
β_3 : 32341	0.518	[0.276]			0.387	[0.278]	0.502	[0.306]

Appendix Chapter 8

This Appendix chapter contains a total of two appendices relevant to Chapter 8, which contain an application of the corrective approach (Appendix A, based on Chapter 7) and details on how to derive a 'loss aversion premium' (Appendix B)

Abbreviations:

TTO	Time trade-off
SG	Standard gamble
QALY	quality-adjusted life year
FH	Full health

Appendix A: Application of the corrective approach

To illustrate the application of the corrective approach on HSV, we use the methods and results presented in Chapter 7. Given that this study used a student sample, yielding weights that were very low compared to the Dutch benchmark tariffs (e.g. Versteegh et al., 2016), we stress that this application only serves illustrative purposes. For this illustration we use simplified notation and 10-year durations for TTO and SG (see Chapter 7 for details). The outcomes in this illustration are denoted as (β, T), where β reflects health status and T duration in years, and we use $\sim$ to indicate indifference, FH to represent full health and D to indicate immediate death. We denote the utility of health status by $U(\beta)$ and the utility of life duration by $L(T)$.

Classical Elicitation Assumption

We consider a typical subject, for whom the following TTO and SG indifferences were elicited (as in Box I): $(\beta, 10) \sim (FH, 6)$ and $(\beta, 10) \sim (FH, 10)_{0.95}(D)$. Although often not stated explicitly, these are the assumptions that apply to the derivations of TTO and SG in tariff estimation for EQ-5D or SF-6D (Versteegh et al., 2016, Walters and Brazier, 2005): a linear QALY model with EU (Pliskin et al., 1980), $U(FH) = 1$ and $U(D) = 0$. Under these assumptions the classical weights for TTO and SG are: $U(\beta_{TTO}) = 0.6$ and $U(\beta_{SG}) = 0.95$.

Corrective Approach

To illustrate the corrective approach, we consider the same typical subject, who is also loss averse ($\lambda = 2$), overweighs extreme probabilities (i.e. $w^+(0.95) = 0.97$, $w^-(1 - 0.95) = 0.10$) and has concave utility of life duration ($L(T) = T^{0.898}$. We will assume that for both TTO and SG the reference point is the time spent in reduced health states (i.e. $(\beta, 10)$). As elaborated in Chapter 7 this allows the following derivations for corrected TTO and SG weights:

$U(\beta_{TTO-C}) = \frac{L(6)-L(0)}{-\lambda(L(6)-L(10))+L(6)-L(0)}$, and $U(\beta_{SG-C}) \frac{w^+(0.95)}{w^+(0.95)+\lambda w^-(1-0.95)}$.

Observation 1: It is straightforward to see that if our subject had behaved as if conforming to the classical elicitation assumption (i.e. $L(T) = T$, $\lambda = 1$, $w^+(p) = p$ & $w^-(1-p) = (1-p)$, the following holds: $U(\beta_{TTO}) = U(\beta_{TTO-C})$ and $U(\beta_{SG}) = U(\beta_{SG-C})$.

Observation 2: If we take into account the subjects' preferences (i.e. loss aversion, probability weighting and concave utility for life duration), we find:

$(\beta_{TTO}) > U(\beta_{TTO-C}) = 0.46$ and $U(\beta_{SG}) > U(\beta_{SG-C}) = 0.83$, i.e. corrected weights are lower.

When this approach was applied for the whole sample, the following classical and corrected weights were elicited in Chapter 7:

		TTO	SG		TTO	SG
Classical	β1: 21211	0.665	0.75	Corrected	0.492	0.506
weight	β2: 31221	0.605	0.706	Weight	0.442	0.456
	β3: 32341	0.39	0.518		0.279	0.319

Appendix B. Loss aversion premium – definition and application to example in Box 8.3

Definition

We would suggest that policy makers should decide, using a deliberative process, whether loss aversion (i.e. distinguishing between health gains and losses) is relevant for the evaluation at hand. If so, a societal loss aversion weight needs to be decided on, which could be factored into utility differences concerning losses. As a first rule of thumb, we suggest a three-category weight: 1) loss neutrality: $\lambda = 1$ (if distinguishing between gains and losses is deemed normatively unacceptable or of negligible societal impact); 2) moderate loss aversion: $\lambda = 1.5$ (Attema et al., 2013a, Attema et al., 2016, Attema et al., 2018); 3) considerable loss aversion: $\lambda = 2$ (Kahneman and Tversky, 1979). The loss aversion premium is subsequently obtained by multiplying the incremental corrected QALY change with the selected loss aversion weight, where correction is necessary to avoid double counting.

Example

For the illustration in Box 8.3 we can find corrected health state weights in Appendix A, where we use the average of the TTO and SG weights: $U(FH) = 1, U(\beta 1) = 0.499, U(\beta 3) = 0.299$. Treatment A results in $U(FH) - U(\beta 1) = 0.501$ and Treatment B has the following difference in utilities: $U(\beta 3) - U(\beta 1) = 0.200$.

Given this, two steps need to be taken in the decision making process if using a premium for losses would be deemed relevant.

Step 1: Decide about sign (i.e. gains or losses)

Treatment A is curative and Treatment B is preventive, i.e. this allocation decision deals with both gains (A) and losses (B). It may be relevant to assign a premium.

Steps 2: Select appropriate loss aversion weight and calculate premium

These final steps rely on judgment by policy makers with regard to whether it is normatively acceptable to assign a premium to losses. Below (in Table O.8.B.1) we show how this categorical decision would impact the utility differences for the illustration in Box 8.3, where the utility difference for Treatment B is multiplied by the selected weight.

Table O.8.B. Loss aversion premium for example in Box 8.3 under different loss aversion weights

Weight	Implication	$\boldsymbol{U(\beta 3) - U(\beta 1)}$	Premium
Loss neutral	$\lambda = 1$	0.200	0
Moderate loss aversion	$\lambda = 1.50$	0.300	0.100
Considerable loss aversion	$\lambda = 2.00$	0.400	0.200

Appendix Chapter 9

This Appendix chapter contains a total of three appendices relevant to Chapter 9, which contain the instructions used and screenshots of the experiment (Appendix A), an elaboration and proofs of the theoretical predictions in Chapter 9 (Appendix B), and additional results to rule out alternative explanations related to acceptability, discounting and exclusion of respondents (Appendix C).

Abbreviations:

TTO	Time trade-off
SG	Standard gamble
QALY	quality-adjusted life year
SLE	subjective life expectancy
FH	Full health

Appendix A: Instructions and choice lists

Below we reprint examples of instructions and choice lists that could be used for a subject who has SLE of 70.

Instructions used for TTO and SG.

Figure O.9.A.1 below demonstrates an example instruction for TTO-gain. For this subject, TTO-losses instruction would be equivalent, except subjects would be instructed to imagine they were 60 years old, and gauge duration would be between age 60 and 70.

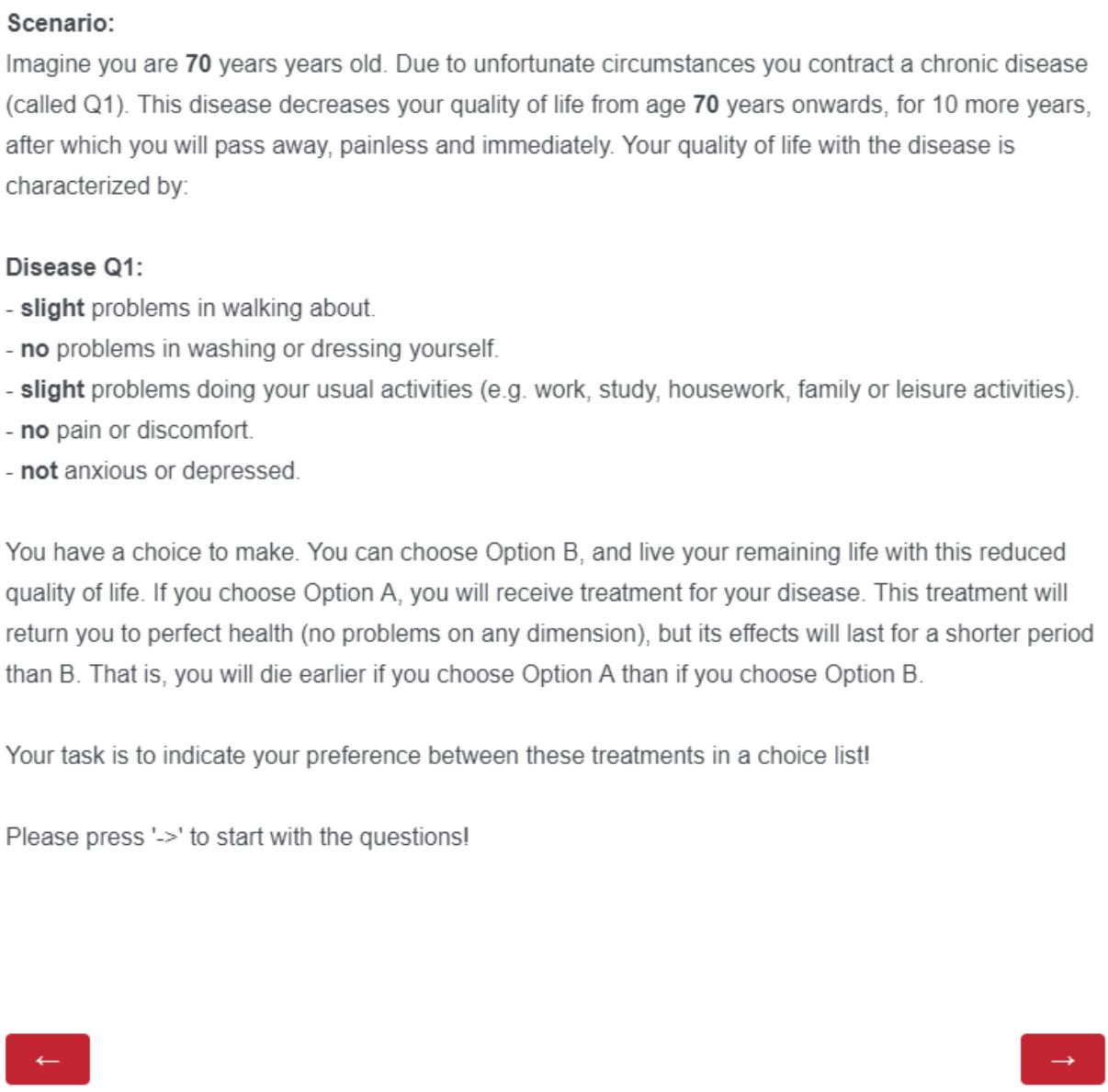


Figure O.9.A.1. Screenshot of instruction screen for TTO-gain

Figure O.9.A.2 shows example instruction for SG-gain. For this subject, SG-losses instruction would be equivalent, except subjects would be instructed to imagine they were **60** years old, and gauge duration would be between age 60 and 70.

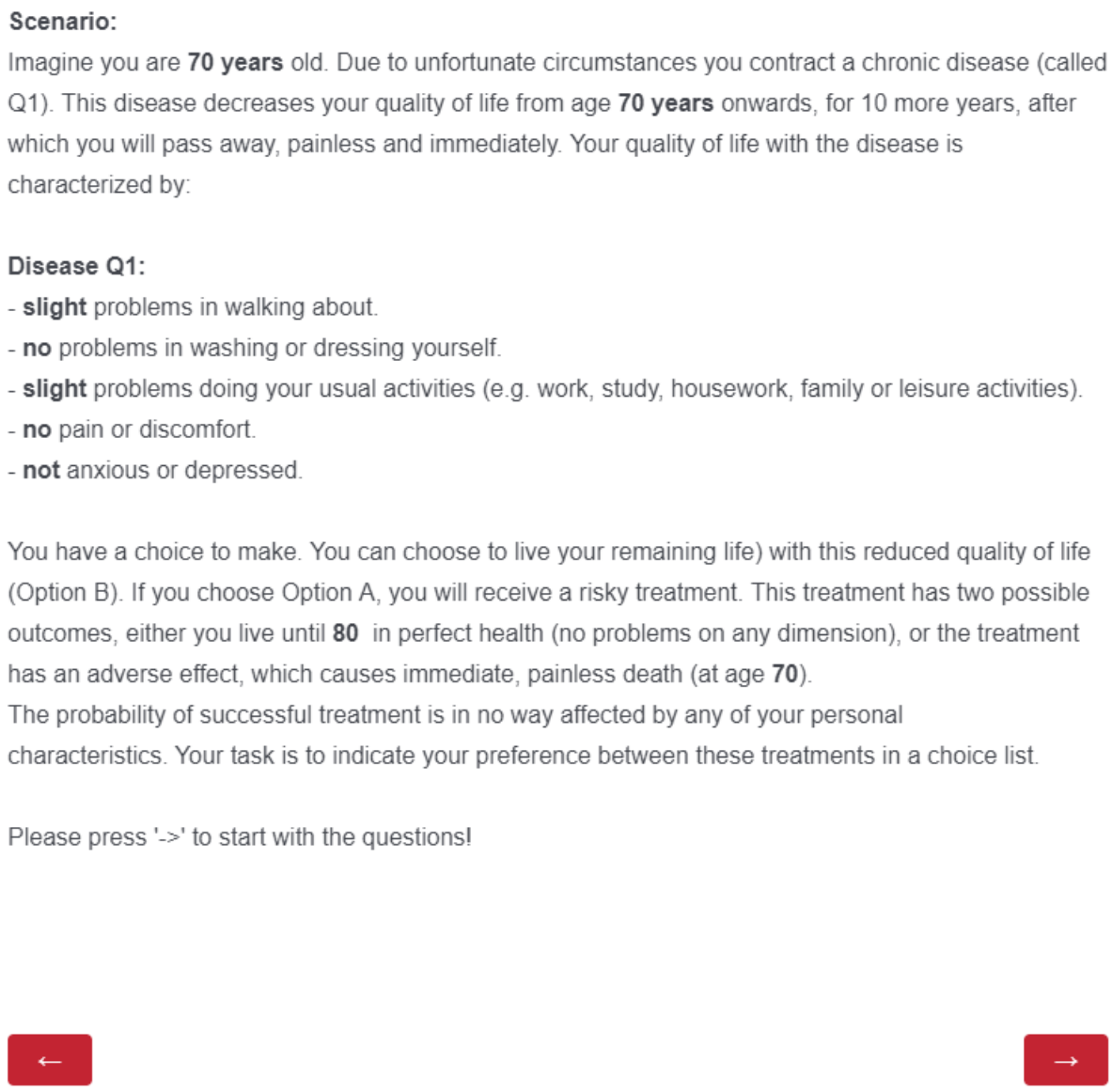


Figure O.9.A.2. Screenshot of instruction screen for SG-gain

Example choice lists.

Figure O.9.A.3 shows an example of first stage choice list, which would be continued with elicitation between 73 and 74 years in monthly increments.

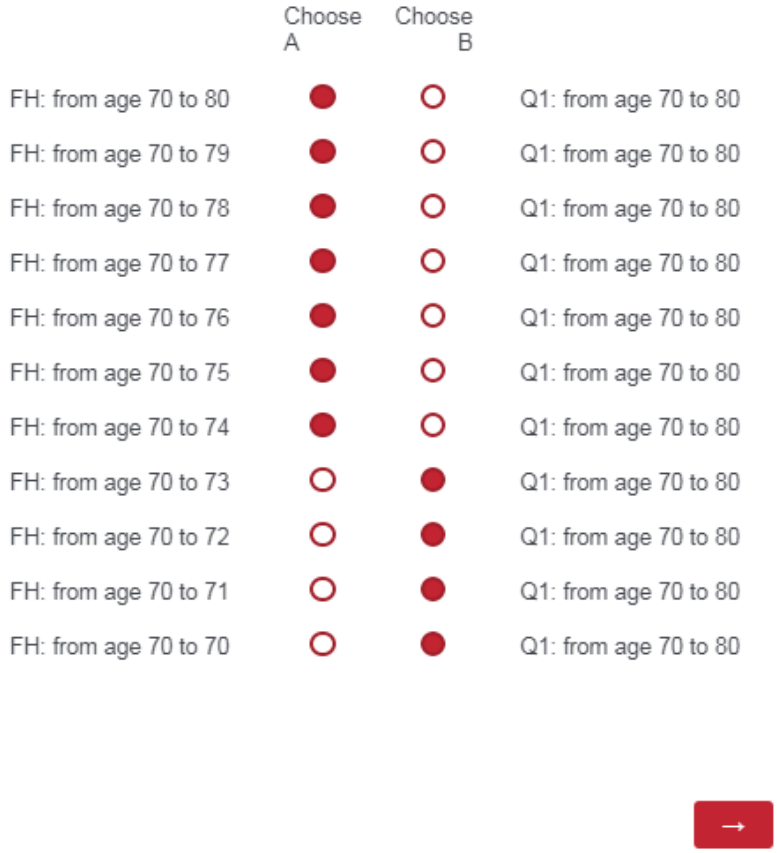


Figure O.9.A.3. Example screenshot for time trade-off choice list

Figure O.9.A.4 shows an example of a first stage choice list for standard gamble, which would be continued with elicitation between 40 and 30% in percentage point increments.

Figure O.9.A.4. Example screenshot for standard gamble choice list

Appendix B: Elaboration and proof for SLE effects

In this Appendix, we apply our theoretical model based on prospect theory to obtain derivations and proofs for both gains and losses versions of TTO and SG. These derivations allow us to illustrate the SLE effect on years given up in TTO and probabilities of immediate death risked in SG. To this end, we will introduce four imaginary respondents, where we make assumptions about the functional forms of their utility function for life duration and probability weighing functions. We provide proof for our predictions by assigning parameter values to these respondents that coincide with the assumptions we made throughout the paper (i.e., loss aversion, S-shaped utility for life duration, and an inverse S-shaped probability weighting function that is more pronounced for gains than losses).

Throughout, we will employ the following structural assumptions. We assume that utility of life duration is captured by a sign-dependent power utility curve, i.e. $L^{+}(T^{*}) = (T^{*})^{\alpha}$ and $L^{-}(T^{*}) = (T^{*})^{\beta}$. Furthermore, we model probability weighting by assuming one-parameter probability weighting (Tversky and Kahneman, 1992), that is:

$$w^{i}(p) = \frac{p^{\gamma}}{(p^{\gamma}+(1-p)^{\gamma})^{1/\gamma}}, \text{ with } i = +, - . \qquad \text{(O9B1)}$$

We will assume that all imaginary respondents have the same SLE, i.e. t_{r} = 70, and all subjects will assign the same utility to health state $Q_x, U(Q_x) = 0.8$.

Consider the following four imaginary respondents:

Respondent A (reference-case): $\alpha, \beta, \gamma, \lambda = 1$,

Respondent B (loss-averse): $\alpha, \beta, \gamma = 1, \lambda = 2$,

Respondent C (sign-dependent): $\alpha, \beta = 0.88, \lambda = 1$ and different γ for gains and losses: $w^{+}(p) = \frac{p^{0.61}}{(p^{0.61}+(1-p)^{0.61})^{1.64}}$ and $w^{-}(p) = \frac{p^{0.69}}{(p^{0.69}+(1-p)^{0.69})^{1.45}}$.

Respondent D (sign-dependent and loss averse): as Respondent C with $\lambda = 2$.

Comparisons between A and B will give the SLE-effect of loss aversion, while comparisons of A and C will show the SLE-effect of sign-dependence of the utility and probability weighting functions. Respondent D represents the case where both SLE-effects are present.

As a reminder, TTO and SG versions for gains and losses are generated by instructing subjects to consider being at different ages throughout the task, i.e. by shifting t_a to t_a^g and t_a^l. As in Section 3, we add superscripts g and l to health status to indicate gains and losses versions of valuation tasks. Throughout, we apply the notation described in section 2 to obtain health profiles in which duration is equated with the reference point, i.e. duration is denoted as T^{*}. For clarity, we give a detailed example of this simplifying notation in Table B1. We also define one additional duration (T_a^{*}) for these derivations. This duration entails living until t_a^g or t_a^l as seen from t_r. Hence, TTO/SG-gains were completed with $t_a^g = t_r$, which in our notation gives $T_a^{*} = 0$ (see Table B1). For TTO/SG-losses we have $t_a^l < t_r$ and our notation gives $T_a^{*} = -T_r$ (see Table O.9.B.1). We add one final notational convention, to let p_g and p_l refer to the probability of immediate death in both SG-gains and SG-losses. That is, we define $p_g, p_l = 1 - p$, giving $\left(\mathrm{Q}_x^g, T_x^{*}\right) \sim (FH^g, T_x^{*})_p(D)$ and

$(\mathrm{Q}_x^l, T_x^*) \sim (FH^l, T_x^*)_p(D)$.

Table O.9.B.1. Notation for a subject who has a SLE of 70 and gives up half of her remaining life time in Q to return to full health in 10 year-TTO for both gains and losses versions

	$Gains\ version, t_a^g = t_r$			
	t_r	t_a^g	(Q_x^g, t_x)	(FH^g, t_y)
t	70	70	80	75
$T = t - t_a^g$	$T_r = 0$	$T_a^g = 0$	$T_x = 10$	$T_y^* = 5$
$T^* = T - T_r$	$T_r^* = 0$	$T_a^* = 0$	$T_x^* = 10$	$T_y^* = 5$
	$Losses\ version, t_a^l = t_r - 10$			
	t_r	t_a^g	(Q_x^g, t_x)	(FH^g, t_y)
t	70	60	70	65
$T = t - t_a^g$	$T_r = 10$	$T_a^l = 0$	$T_x = 10$	$T_y = 5$
$T^* = T - T_r$	$T_r^* = 0$	$T_a^* = -10$	$T_x^* = 0$	$T_y^* = -5$

Formal derivations for years given up in TTO

TTO-gains $(t_a^g = t_r)$: $T_x^* \geq T_y^* \geq T_a^* = 0$

TTO indifferences of the form $\left(\mathrm{Q}_x^g, T_x^*\right) \sim (FH^g, T_y^*)$ in this version are evaluated depending on the distance between T_y^* and SLE (T_r). As long as all considered life years are spent above SLE ($T_x^* \geq T_y^* \geq 0$), all life years are above the RP and can thus be considered gains. Under these conditions, $\left(\mathrm{Q}_x^g, T_x^*\right) \sim (FH^g, T_y^*)$ can be evaluated by:

$$L^+(T_x^*)U\left(\mathrm{Q}_x^g\right) = L^+\left(T_y^*\right)U(FH^g). \quad \text{(O9B2)}$$

With $U(FH^g) = 1$, solving for T_y^* gives:

$$T_y^* = {L^+}^{-1}\left(L^+(T_x^*)U(Q_x)\right). \quad \text{(O9B3)}$$

Observation 1.1.a: Respondent A and B will have $T_y^* = 10 \times 0.8 = 8$.

Observation 1.1.b: Respondent C and D will have $T_y^* = (10^{0.88} \times 0.8)^{1/0.88} = 7.76$.

TTO-losses ($t_a^l < t_r$): $T_a^* < T_y^* < T_x^* = 0$

In the loss version, the life years will occur completely below SLE, and if SLE functions as RP these life years are considered to be losses. Thus, TTO indifferences of the form $(Q_x^l, T_x^*) \sim (FH^l, T_y^*)$ entail indifference between a loss in life years of $-T_y^*$ years with an improvement in health status experienced for $T_y^* - T_a^*$ years, or living up to SLE (i.e. $T_x = T_r,\ T_x^* = 0$). Under these conditions, $(Q_x^l, T_x^*) \sim (FH^l, T_y^*)$ can be evaluated by:

$$L^+(T_x^*)U(Q_x^l) = \lambda L^-(T_y^*)U(Q_x^l) + \left(U(FH^g) - U(Q_x^l)\right)\left[L^-(T_y^*) - L^-(T_a^*)\right]. \quad \text{(O9B4)}$$

Because $L^+(T_x^*) = L^+(0) = (0)^\alpha = 0$, and $U(FH^g) = 1$, this expression simplifies into:

$$0 = \lambda L^-(T_y^*)U(Q_x^l) + \left(1 - U(Q_x^l)\right)\left[L^-(T_y^*) - L^-(T_a^*)\right]. \quad \text{(O9B5)}$$

Rearranging yields:

$$\lambda L^-(T_y^*)U(Q_x^l) + \left(1 - U(Q_x^l)\right)\left[L^-(T_y^*)\right] = \left(1 - U(Q_x^l)\right)L^-(T_a^*). \quad \text{(O9B6)}$$

Solving for T_y^* gives:

$$T_y^* = L^{--1}\left(\frac{\left(1-U(Q_x^l)\right)L^-(T_a^*)}{U(Q_x^l)(\lambda-1)+1}\right). \quad \text{(O9B7)}$$

Observation 1.2.a: Respondent A will have $T_y^* = \left(\frac{(1-0.8)\times-10}{1}\right) = -2.$

Observation 1.2.b: Respondent B will have

$$T_y^* = L^{--1}\left(\frac{(1-0.8)\times-10}{0.8+1}\right) = -\frac{2}{1.8} = -1.11.$$

Observation 1.2.c: Respondent C will have $T_y^* = -\left(-\frac{(1-0.8)\times(-10)^{0.88}}{1}\right)^{1/0.88} = -1.61.$

Observation 1.2.d: Respondent D will have $T_y^* = -\left(-\frac{(1-0.8)\times(-10)^{0.88}}{1.8}\right)^{1/0.88} = -0.82.$

Comparison between TTO-gains and TTO-losses

To compare the years given up between the gains and losses version of TTO, we can compare answers by Respondents A to D across versions. For ease of comparability, Table O.9.B.2 compares respondents A to D for gains and losses under our various forms of notation (as in Table B1). As can be seen from Table 2 (in Chapter 9), the effects predicted in our theoretical model occur. That is, in the reference case, the same amount of years is given up between versions. Furthermore, compared to the reference case (A), incorporating loss aversion (B) yields fewer life years given up. Similarly, compared to the reference case (A), incorporating a sign-dependent utility curve (C) yields fewer life years given up. Combining both influences (D) yields even fewer life years given up. These effects should give higher QALY weights for TTO-losses.

Table O.9.B.2. T_y^* and T_y for TTO-gains and TTO-losses versions for respondents A to D.

Version	Gain	Loss		Gain	Loss	
	(FH^g, T_y^*)	(FH^l, T_y^*)		(FH^g, T_y)	(FH^l, T_y)	Diff. in years
Respondent A	$T_y^* = 8$	$T_y^* = -2$		$T_y = 8$	$T_y = 8$	0
Respondent B	$T_y^* = 8$	$T_y^* = -1.11$		$T_y = 8$	$T_y = 8.89$	-0.89
Respondent C	$T_y^* = 7.76$	$T_y^* = -1.61$		$T_y = 7.76$	$T_y = 8.39$	-0.63
Respondent D	$T_y^* = 7.76$	$T_y^* = -0.82$		$T_y = 7.76$	$T_y = 9.18$	-1.42

Formal derivations of probability of D accepted for SG

SG-gains ($\boldsymbol{t_a^g = t_r}$): $\boldsymbol{T_x^* \geq T_a^* = 0}$

In this version, all outcomes are at or beyond the SLE, i.e. all life years are considered to be gains. SG indifferences of the form $\left(\mathrm{Q}_x^g, T_x^*\right) \sim (FH^g, T_x^*)_p(D)$ consist of indifferences between a sure gain in life years of T_x^* years, and an uncertain gain, yielding either an improvement in quality of life for the same T_x^* years, or immediate death T_r, which implies living until t_a^g. When $t_a^g = t_r$, D entails staying at T_r, i.e. $T_a^* = 0$. For consistency we will equate these 0 years with $U\left(\mathrm{Q}_x^g\right)$, and hence this indifference can be evaluated by:

$$U\left(\mathrm{Q}_x^g\right)L^+(T_x^*) = w^+(p)\left[(U(FH^g) - U\left(\mathrm{Q}_x^g\right))L^+(T_x^*) + U\left(\mathrm{Q}_x^g\right)L^+(T_x^*)\right] + \left(1 - w^+(p)\right)U\left(\mathrm{Q}_x^g\right)L^+(T_a^*). \quad \text{(O9B8)}$$

If we set $U(FH^g) = 1$ and $T_a^* = 0$, this expression becomes:

$$U\left(\mathrm{Q}_x^g\right)L^+(T_x^*) = w^+(p)\left[(1 - U\left(\mathrm{Q}_x^g\right))L^+(T_x^*) + U(Q_x)L^+(T_x^*)\right] + \left(1 - w^+(p)\right)U\left(\mathrm{Q}_x^g\right)L^+(0), \quad \text{(B9)}$$

which can be simplified into:

$$U\left(\mathrm{Q}_x^g\right)L^+(T_x^*) = w^+(p)L^+(T_x^*) \leftrightarrow w^+(p) = U\left(\mathrm{Q}_x^g\right). \quad \text{(O9B10)}$$

Solving for p yields:

$$p = (w^+(p))^{-1}U\left(\mathrm{Q}_x^g\right), \quad \text{(O9B11)}$$

and hence:

$$p_g = 1 - (w^+(p))^{-1}\left(U\left(\mathrm{Q}_x^g\right)\right). \quad \text{(O9B12)}$$

Observation 2.1.a: Respondent A and B will have $p_g = 0.2$.

Observation 2.1.b: Respondent C and D will have $p_g = 0.08$.

SG-losses ($t_a^l < t_r$): $\boldsymbol{T_a^* < T_x^* = 0}$

SG indifferences of the form $(\mathrm{Q}_x^l, T_x^*) \sim (FH^l, T_x^*)_p(D)$ in the loss scenario consist of indifferences between living up to t_{SLE} in Q_x^l, and a risky outcome that either yields an improvement in quality of life for the same duration T_x^*, or a loss of T_x^* years (immediate death). This can be evaluated as follows:

$$U(\mathrm{Q}_x^l)L^+(T_x^*) = w^+(p)(U(FH^l) - U(\mathrm{Q}_x^l))\left(L^+(T_x^*) - L^-(T_a^*)\right) + \lambda w^-(1-p)U(\mathrm{Q}_x^l)L^-(T_a^*). \quad \text{(O9B13)}$$

By setting set $U(FH^l) = 1$ and $T_x^* = 0$, we obtain:

$$U(\mathrm{Q}_x^l)L^+(0) = w^+(p)\left(1 - U(\mathrm{Q}_x^l)\right) \times \left(-L^-(T_a^*)\right) + \lambda w^-(1-p)U(\mathrm{Q}_x^l)L^-(T_a^*). \quad \text{(O9B14)}$$

Rearranging gives:

$$0 = w^+(p)(U(\mathrm{Q}_x^l) - 1) \times \left(L^-(T_a^*)\right) + \lambda w^-(1-p)U(\mathrm{Q}_x^l)L^-(T_a^*) \leftrightarrow$$

$$w^+(p)\left(1 - U(\mathrm{Q}_x^l)\right) = \lambda w^-(1-p)U(\mathrm{Q}_x^l) \leftrightarrow$$

$$\frac{w^+(p)}{w^-(1-p)} = \frac{\lambda U(\mathrm{Q}_x^l)}{1-U(\mathrm{Q}_x^l)}. \quad \text{(O9B15)}$$

Since solving this equation analytically for p gets somewhat complicated and is beyond the purpose of this Appendix (which is to compare the predicted probability of D accepted by our imaginary respondents given in the different versions) we use a numerical approach to derive p_l for each respondent.

Observation 2.2.a: Respondent A will have $\frac{p}{1-p} = \frac{0.8}{0.2} \leftrightarrow p = 0.8, p_l = 0.2$.

Observation 2.2.b: Respondent B will have $\frac{p}{1-p} = \frac{2\times 0.8}{0.2} \leftrightarrow p = 0.89, p_l = 0.11$.

Observation 2.3.b: Respondent C will have $\frac{\frac{p^{0.61}}{\left(p^{0.61}+(1-p)^{0.61}\right)^{1/0.61}}}{\frac{(1-p)^{0.69}}{\left((1-p)^{0.69}+(p)^{0.69}\right)^{1/0.69}}} = \frac{0.8}{0.2} \leftrightarrow p = 0.894, p_l = 0.106$.

Observation 2.4.b: Respondent D will have $\frac{\frac{p^{0.61}}{\left(p^{0.61}+(1-p)^{0.61}\right)^{1/0.61}}}{\frac{(1-p)^{0.69}}{\left((1-p)^{0.69}+(p)^{0.69}\right)^{1/0.69}}} = \frac{1.6}{0.2} \leftrightarrow p = 0.9648, p_l = 0.035$.

Comparison between SG-gains and SG-losses

As can be seen from Table O.9.B.3, the effects predicted in by our theoretical model occur (at least in Study 1. That is, in the reference case (A), the same probability of death is accepted.

Furthermore, compared to the reference case (A), incorporating loss aversion (B) yields a lower probability of D to be accepted. In contrast, compared to the reference case (A), incorporating sign-dependent probability weighting (C) yields a lower probability of D to be accepted for losses versions. Combining both influences (D) yields an even lower probability of D to be accepted for losses versions, although the exact size and direction of this combined effect will depend on the parameters selected for λ and γ for gains and losses.

Table O.9.B.3: Comparisons of probabilities of D accepted in SG-gains and SG-losses

Version	Gain	Loss	Diff.
$p_g, p_l = 1 - p$	$\left(\mathrm{Q}_x^g, T_x^*\right) \sim (FH^g, T_x^*)_p(D)$	$(\mathrm{Q}_x^l, T_x^*) \sim (FH^l, T_x^*)_p(D)$	
Respondent A	$p_g = 0.2$	$p_l = 0.2$	0
Respondent B	$p_g = 0.2$	$p_l = 0.11$	-0.09
Respondent C	$p_g = 0.08$	$p_l = 0.106$	0.026
Respondent D	$p_g = 0.08$	$p_l = 0.035$	-0.045

Appendix C: Manipulation checks for acceptability of health states and discounting

In this Appendix we aim to demonstrate that our findings are not related to acceptability of health states, exclusion rates and discounting (from current age).

Acceptability of health states

First, we created a correlation matrix to test if our manipulation was successful and if any of the residual reference outcomes affected QALY weights for our TTO and SG estimates (see Table O.9.C.1 for Study 1 and Table O.9.C.2 for Study 2). Perhaps unsurprisingly, the reference outcomes (1 to 6) are in many cases correlated positively and significantly. Importantly, SLE is only correlated with QALY weights in one condition (without Bonferroni correction). This suggests that no systematic association for SLE appears to remain after our experimental manipulation, and violations of our restriction to acceptable health states did not affect TTO and SG responses. Interestingly, we find systematic and significant positive correlations between the maximum age individuals are willing to attain and QALY weights for all conditions in Study 2. This finding suggests that those who are willing to live to a higher age give up fewer life years in TTO and risk lower chances of immediate death in SG. No such associations are found for the minimum age people want to become or their SLE.

Additionally, we explored the effect of acceptability on our main results. First, for each completed TTO or SG, we determined if the health state was deemed acceptable at that age, which was categorized in a binary variable. Next, we ran mixed effects regression models (with R package lmer) on the normalized QALY weights (i.e. after applying Eq. 9.2 and Eq. 9.3). The model included subject random effects, and fixed effects for: a) method (TTO vs. SG), b) version (Gain vs. Loss), c) health state (Q1, Q2, and Q3) and acceptability (Not-Acceptable vs. Acceptable), and d) sex (male vs. female). We also included an interaction term for method and version. In Study 2 we also include having chronic disease as a fixed effect. These analyses (Table O.9.C.3 and Table O.9.C 4) indicate that our main result, i.e. higher utilities for loss versions, which was more pronounced for TTO, remains intact and no effect of acceptability (or sex) is observed. Having a chronic disease does affect TTO and SG weights.

Taken together, these two analyses give no direct reason to expect that acceptability of health status affected our main results. As such, even though we restricted our theoretical model to cases where health status is better than accepted, we decided to retain the full sample, as violations of this assumption did not appear to affect any of our results. A similar conclusion applies to the experience of chronic disease in Study 2.

Table O.9.C.1. Correlation matrix for subjective life expectancy (SLE), acceptability of health states (Q1, Q2 and Q3), time trade-off (TTO) and standard gamble (SG) weights in Study 1.

	(1)	(2)	(3)	(4)	(5)	(6)
1: SLE	-					
2: SLE-min	0.36**	-				
3: SLE-max	0.40***	0.32**	-			
4: Acceptable age (Q1)	0.09	0.37**	0.25*	-		
5: Acceptable age (Q2)	0.22	0.47***	0.31**	0.82***	-	
6: Acceptable age (Q3)	0.24*	0.50***	0.37**	0.70***	0.91***	-
Q1 – TTO - gain	0.04	0.18	0.07	0.14	0.14	0.12
Q1 – TTO - loss	0.01	0.05	0.08	0.05	0.04	0.02
Q2 – TTO - gain	-0.02	-0.04	-0.04	0.05	0.12	0.10
Q2 – TTO - loss	-0.28**	0.01	-0.10	0.13	0.14	0.10
Q3 – TTO - gain	-0.09	0.03	-0.09	-0.02	0.01	-0.03
Q3 – TTO - loss	-0.01	0.02	-0.06	0.05	0.05	0.01
Q1 – SG - gain	-0.18	0.10	-0.08	0.05	-0.05	-0.07
Q1 – SG - loss	-0.17	0.08	-0.13	0.05	0.01	-0.05
Q2 – SG - gain	-0.11	0.05	-0.08	0.03	0.00	-0.05
Q2 – SG - loss	0.06	-0.02	0.04	-0.14	-0.11	-0.11
Q3 – SG - gain	0.04	-0.03	0.12	-0.07	-0.01	0.02
Q3 – SG - loss	0.05	0.09	0.15	-0.02	0.01	0.01

Note: *,** and *** indicate significance at $p = 0.05$, $p = 0.01$ and $p = 0.001$ respectively

Table O.9.C.2. Correlation matrix for subjective life expectancy (SLE), acceptability of health states (Q1, Q2 and Q3), time trade-off (TTO) and standard gamble (SG) weights in Study 2.

	(1)	(2)	(3)	(4)	(5)	(6)
1: SLE						
2: SLE-min	0.48***					
3: SLE-max	0.47***	0.60***				
4: Acceptable age (Q1)	0.20***	0.28***	0.25***			
5: Acceptable age (Q2)	0.23***	0.29***	0.27***	0.91***		
6: Acceptable age (Q3)	0.26***	0.32***	0.36***	0.81***	0.92***	
Q1 – TTO - gain	0.02	0.13*	0.18**	0.02	0	0.02
Q1 – TTO - loss	0.09	0.05	0.13*	-0.03	-0.05	-0.01
Q2 – TTO - gain	0.03	0.14*	0.22***	-0.05	-0.07	-0.03
Q2 – TTO - loss	0.06	0.13*	0.22***	-0.06	-0.08	-0.03
Q3 – TTO - gain	-0.05	0.03	0.15**	-0.09	-0.16**	-0.13*
Q3 – TTO - loss	0.01	0.15**	0.16**	-0.06	-0.11	-0.08
Q1 – SG - gain	0.05	0.08	0.20***	-0.02	-0.02	0.01
Q1 – SG - loss	0.06	0.05	0.21***	0	0.01	0.04
Q2 – SG - gain	0.03	0.05	0.18***	-0.03	-0.04	-0.01
Q2 – SG - loss	0.05	0.05	0.17**	-0.04	-0.04	-0.02
Q3 – SG - gain	0.05	0.1	0.19***	-0.07	-0.09	-0.06
Q3 – SG - loss	0.07	0.12*	0.22***	-0.04	-0.05	-0.04

Note: *,** and *** indicate significance at $p = 0.05$, $p = 0.01$ and $p = 0.001$ respectively

Table O.9.C.3. Fixed effect estimates (standard errors) for mixed effect regression analyses on TTO and SG weights for Study 1 (students).

	Estimate	*SE*	*df*	*T*	*p*
Constant	0.56	0.018	117	12.81	**<0.001**
Method: SG	0.20	0.015	1116	12.92	**<0.001**
Version: Loss	0.15	0.015	1137	9.30	**<0.001**
Health state: Q2	-0.06	0.013	1118	-4.51	**<0.001**
Health state: Q3	-0.18	0.013	1157	-12.83	**<0.001**
Method*Version	-0.10	0.022	1117	-4.63	**<0.001**
Acceptability: Yes	0.00	0.020	1160	0.15	0.88
Gender: Female	0.02	0.030	100	0.50	0.61

Table O.9.C.4. Fixed effect estimates (standard errors) for mixed effect regression analyses on TTO and SG weights for Study 2 (respondents age 60 years and older).

	Estimate	*SE*	*df*	*T*	*p*
Constant	0.64	0.017	494	38.35	**<0.001**
Method: SG	0.15	0.009	3549	15.99	**<0.001**
Version: Loss	0.11	0.010	3666	10.64	**<0.001**
Health state: Q2	-0.02	0.008	3572	-2.74	**0.006**
Health state: Q3	-0.09	0.009	3648	-10.42	**<0.001**
Method*Version	-0.09	0.013	3549	-6.76	**<0.001**
Acceptability: Yes	-0.01	0.020	3855	-0.97	0.33
Gender: Male	-0.02	0.030	320	-0.73	0.47
Chronic disease: Yes	0.05	0.020	320	2.31	**0.02**

Main results for several subsamples for respondents aged 60 years and older (Study 2)

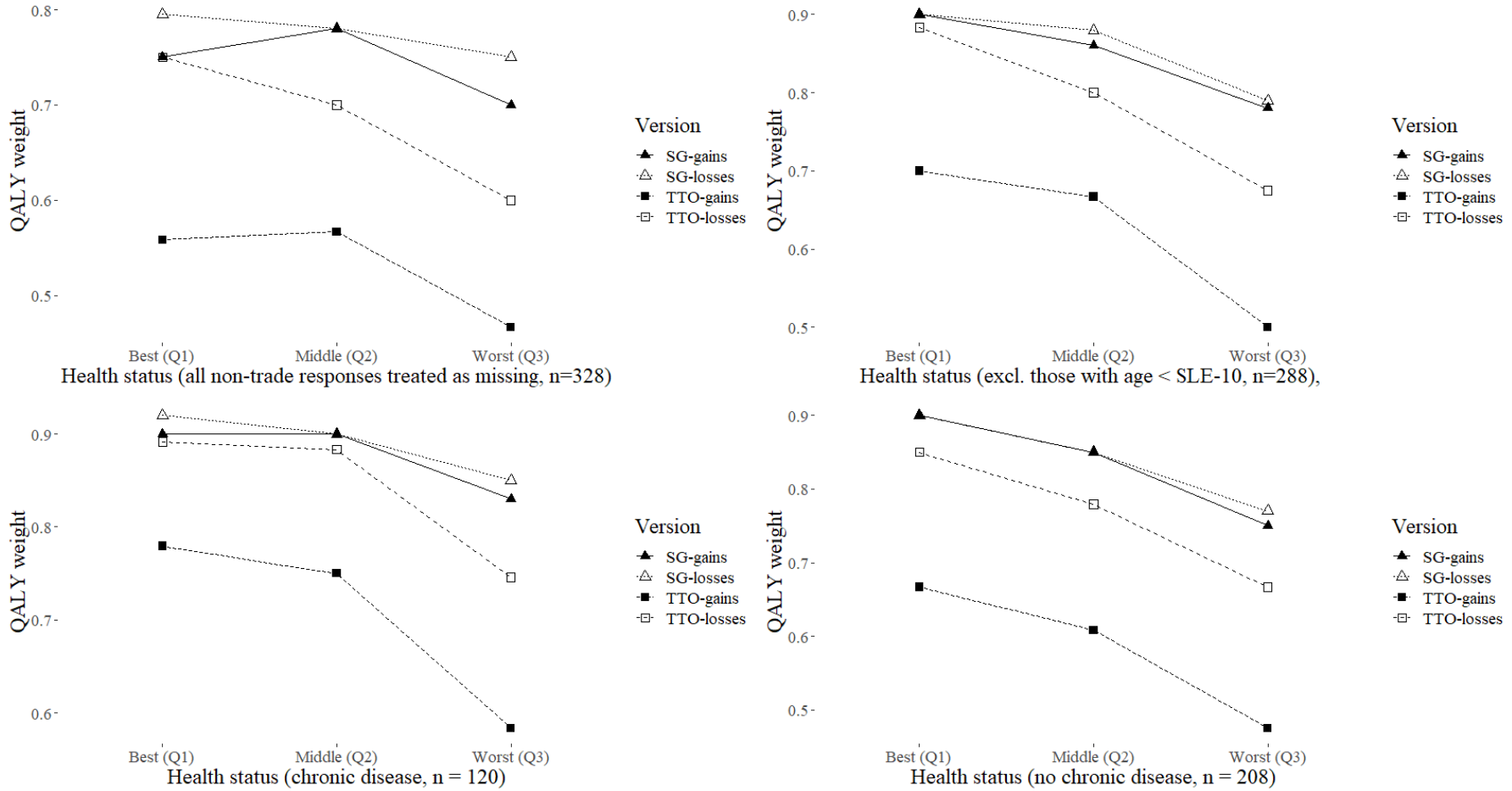


Figure O.9.C.1. Figure C1: Study 2 main results (i.e. Figure 4) for different groups and exclusion rules (upper left: without non-trading responses, upper right: without those for whom t_a^l was lower than their current age, lower left: those with chronic disease, lower right: those without chronic disease)

Discounting (from current age) for TTO

As a final manipulation check, we show that discounting has no or a very small effect on responses for subjects who do not answer according to the instructions but instead erroneously assume their t_a corresponds to their current age (denoted t_a^C, i.e. subjects do not take into account the shift from t_a^g to t_a^l). As a result, health profiles (Q, t) will have durations T of $t - t_a^C$ years. First, we show that in case of constant (exponential) discounting, there is no effect, because the longer delay will affect the gain and loss frames equally. Second, we show that in case of power discounting, there is a small effect because of the non-constant discount rate implicit in power discounting, which means that the exact amount of delay will impact the difference between the gain and loss frames. However, as becomes clear from the exercise below, this effect is small and in the opposite direction of what we found, and, hence, this possibility cannot explain our results.

Consider a subject of 20 years old, with an SLE of 80. Hence, for this subject the TTO-gains has t_a^g = 80 and TTO-losses starts at $t_a^l = 70$. TTO-gains will involve eliciting full-health equivalents for living from age 80 until 90 in Q_x, while TTO-losses versions will involve eliciting full-health equivalents for living from age 70 until 80 in Q_x. In both cases, the duration of the time spent in Q_x is 10 years. Now, if she mistakenly takes t_a to be t_a^C, the gain frame duration T_x = 70 years (i.e. 90-20), whereas the loss frame will have a duration of T_x = 60 years (i.e. 80-20). We derive the predicted answers given by this subject, assuming her utility of this health state $U(Q_x) = 0.5$ and she is not affected by sign-dependence. We will make one final simplifying assumption: the subject considers the years not accounted for by TTO (i.e. before onset of Q_x) are spent in full health.

We start with the case of a constant discounter with a discount rate (r) of 5% per annum. The subject's indifference in TTO-gains can be evaluated by:

$$\frac{1-1.05^{-60}}{0.05} + \left(\frac{1-1.05^{-70}}{0.05} - \frac{1-1.05^{-60}}{0.05}\right) \times 0.5 = \frac{1-1.05^{-X}}{0.05}. \qquad \text{(C1)}$$

Solving for X gives X=64.40. Hence, this subject will be indifferent between living 4.40 years in full health from age 80, and living in heath state Q_x from age 80 till 90, yielding a TTO weight of 4.40/10=0.44 when Eq. 2 is applied (i.e. assuming linear QALYs).

For TTO-losses, we can replace 60 by 50 and 70 by 60 in Eq. C1, to yield X=54.40, again resulting in a TTO weight of 0.44 when Eq. 2 is applied. It is straightforward to show that this result holds for any value of r and t_a^C.

Let us continue with the same example for a power discounter with α=0.7 and all other parameters unchanged. Now, the subject's indifference for TTO-gains, with $U(FH^g) = 1$, is evaluated by:

$$60^{0.7} + (70^{0.7} - 60^{0.7}) \times 0.5 = X^{0.7}. \qquad \text{(C2)}$$

Solving for X gives X=64.94, for a TTO weight of 0.494. Similarly, for TTO-losses we obtain X=54.93 and a TTO weight of 0.493. Hence, the TTO weight decreases just slightly, but negligibly as opposed to the magnitude of the effects observed in Table 3 (or Table B2 in Appendix B).

Appendix Chapter 10

This Appendix chapter contains a total of three appendices relevant to Chapter 10, which contain the instructions used (Appendix A), screenshots (Appendix B), and additional results on the effect of collective decision-making on precision, completion times and bargaining weights (Appendix C).

Abbreviations:

TTO	Time trade-off
SG	Standard gamble
QALY	quality-adjusted life year
IDM	individual decision-making
CDM	collective decision-making

Appendix A: Example instructions

The instructions below were used for the tasks in Chapter 10.

In part 1, you have to perform 2 tasks.

Task 1 (standard gamble)

Suppose you have to choose between 2 possible life scenarios, which are referred to as Alternative A and Alternative B. In Alternative A, you will be certain to live 50 more years in the indicated health state, after which you will die. For example, suppose the health state is as given below:

Your health state (P):

-You have severe problems in walking about
-You have no problems in washing or dressing yourself
-You have moderate problems doing your usual activities (e.g. work, study, housework, family or leisure activities)
-You have slight pain or discomfort
-You are not anxious or depressed

If you choose Alternative B, you are taking a gamble. On the one hand, you have the chance (X%) of living 50 more years in full health (i.e. no problems on any dimension), after which you will die, but on the other hand, you have a chance (100-X %) of dying within a week.

The task consists of a number of lists of choices between the two alternatives. In every list, Alternative A remains the same, but Alternative B varies.

As you move down the list, Alternative B becomes more attractive, and in some row, you will probably switch from Alternative A to Alternative B. If so, you will also choose Alternative B in all rows below that one, because in these Alternative B is more attractive. Similarly, if you choose Alternative A in a given row, you will also choose Alternative A in all rows above that one, because in these Alternative B is less attractive. The computer takes this into account and automatically selects Alternative B for all rows below the one where you choose Alternative B and Alternative A for all rows above the one where you choose Alternative A.There are no right or wrong answers, we are only interested in your choices.

You can change your choices as often as you like. Once you are satisfied with your choices, click the “OK” button. Then you can no longer change your choices and you receive the next choice list. Please now choose the alternative you prefer in each row. If you are ready, you get a prompt on your screen. At that moment, please read the instruction of Task 2 on the next page.

Task 2 (time trade-off)

Again, suppose you have to choose between 2 possible life scenarios, which are referred to as Alternative A and Alternative B. In Alternative A, you will live 50 more years in the indicated health state, after which you will die. For example, suppose the health state is as given below:

Your health state (P):

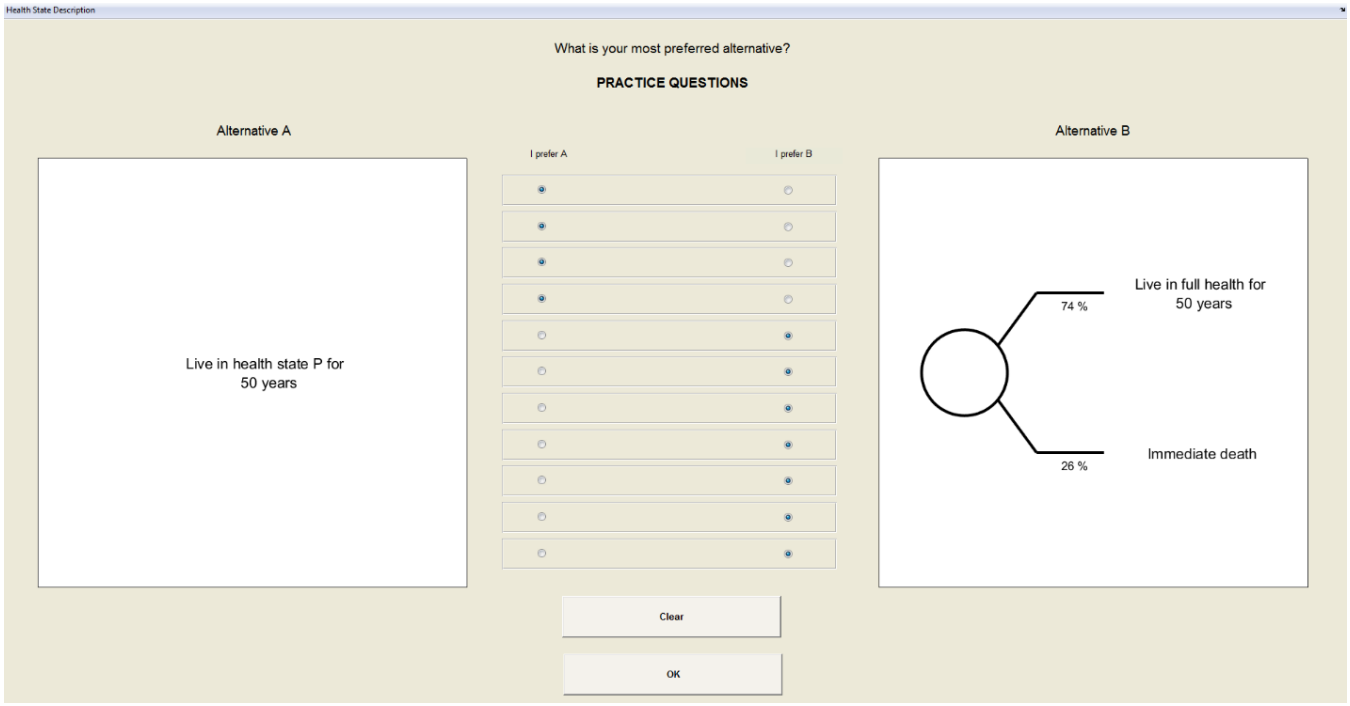
-You have severe problems in walking about
-You have no problems in washing or dressing yourself
-You have moderate problems doing your usual activities (e.g. work, study, housework, family or leisure activities)
-You have slight pain or discomfort
-You are not anxious or depressed

If you choose Alternative B, you will live X more years in full health (i.e. no problems on any dimension), after which you will die.

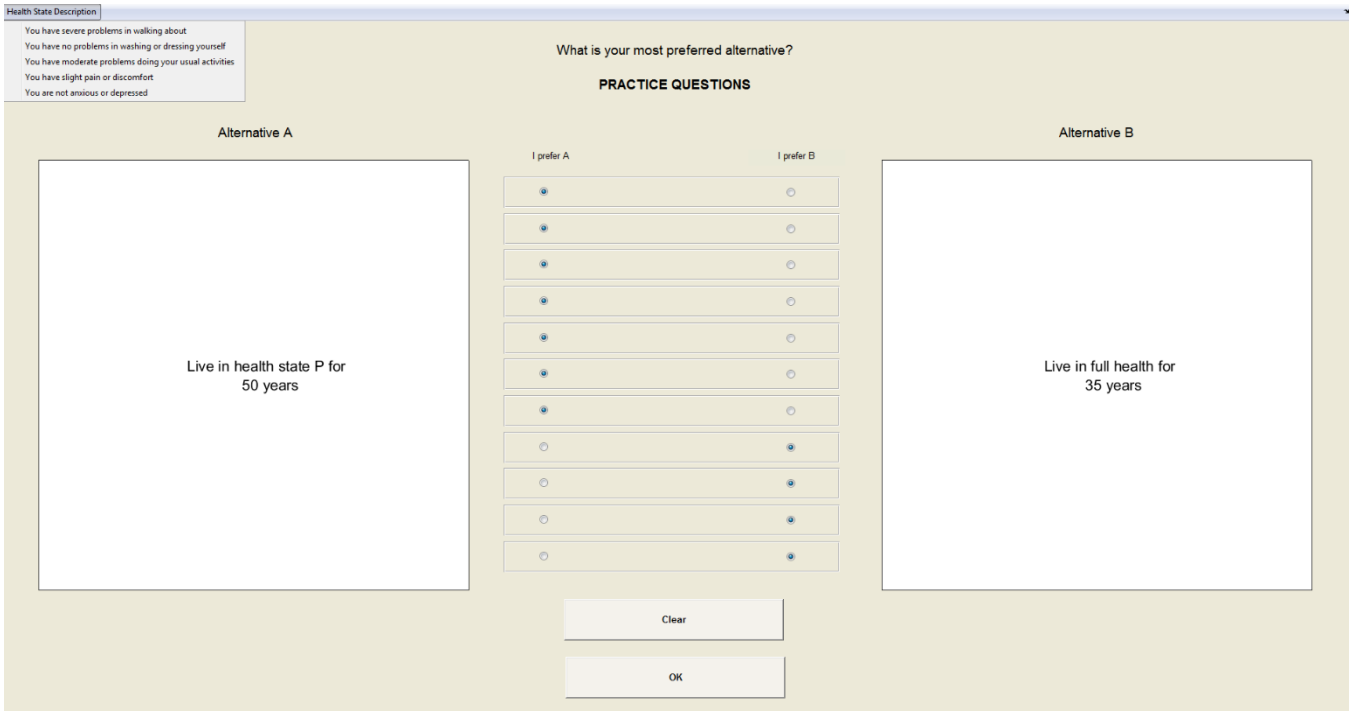
Please choose the alternative you prefer in each row. This procedure is similar as in Task 1.

Appendix B: Screenshots of the experimental program

Figures O.10.B.1 and O.10.B.2 illustrate how the experimental program looked like.



Figures O.10.B.1. Example screenshot for task 1 standard gamble



Figures O.10.B.1. Example screenshot for task 2 (time trade-off)

Appendix C: Additional results on precision, completion time and decision process for SG and TTO

Two additional elements of quality of decision making were analyzed, the precision of utility weights and the completion time for each elicitation. We also estimated the group and carryover effect for these decision elements, which can be found in Table O.10.C.1.

Table O.10.C.1. Fixed effect estimates (standard errors) for LMER analyses for both group and carryover effects

	Decision process	
	Precision	**Time**
Group effect : IDM: I1 vs. I2 | CDM: I1 vs G		
Constant	0.03 (0.01) ***	72.74 (3.86) ***
Learning	-0.004 (0.004)	-15.91 (1.99) ***
Treatment	-0.005 (0.009)	-17.37 (4.75) ***
Method: TTO	0.01 (0.002)***	-12.91 (1.25) ***
Group: (Learning*Treatment)	0.009 (0.005) +	15.32 (2.55) ***
Health state: middle		5.54 (1.53) ***
Health state: high		13.55 (1.53) ***
Group effect : IDM: I1 vs. I2 | CDM: I1 vs G		
Constant	0.03 (0.01) ***	76.54 (2.65) ***
Learning	-0.004 (0.004)	-19.89 (1.74) ***
Treatment	-0.001 (0.009)	4.49 (4.11)
Method: TTO	0.01 (0.003) ***	-13.65 (1.09) ***
Carryover (Learning*Treatment)	0.006 (0.005)	-6.53 (2.24) **
Health state: middle		6.34 (1.34) ***
Health state: high		20.35 (1.34) ***

Note: *,**, and *** represent significance at $p < 0.05$, 0.01 and 0.001 respectively. + indicates marginal significance at $0.05 < p < 0.10$.

Precision

Precision was analyzed both between-subjects and within-subjects. For between-subjects comparisons, we apply Morgan-Pittman tests for equality of variances to compare between

session variance within-methods. For example, we compare SG weight variance for state Q_1 between session I1 and session I2. These tests indicated the degree to utility weights were heterogeneous between sessions and health states. For IDM variances were not significantly different between I1 and I2 (Morgan-Pittman tests, all p's > 0.16). If we repeat these analyses (I1 vs I2) for CDM, we find a significant decrease (Morgan-Pittman tests, all p's < 0.034) in variance, with the exception of the most severe health state Q_3 for both SG and TTO (Morgan-Pittman tests, p's > 0.15). For CDM, we observe significantly smaller variance between the first individual session and group task (Morgan-Pittman tests, all p's < 0.023). The estimation of fixed group or carryover effects is not possible, as these variance estimates reflect between-subjects heterogeneity. Second, we obtain within-subjects estimates of precision by calculation of variance for utility weights associated with Q_1, Q_2 and Q_3 (see Table O.10.C.2). These analyses indicate to what extent collective decision making affected dispersion of utility weights for each individual, i.e. if utility weights elicited in each session become more condensed or dispersed. Next, when we applied our analytical approach to estimate for the group effect and carryover on within-subject variance (see Table C1), we observed only a fixed effect of method, implying higher dispersion for TTO compared to SG. We observed no effects of learning, treatment, group or carryover effects of collective decision making.

Table O.10.C.2. Decision quality: Mean within-subjects variance and percentages of subjects satisfying monotonicity for each session

	Session 1		Session 2		Session 3
	I1-IDM	I1-CDM	I2-IDM	Group	I2-CDM
Variance for $Q_1, Q_2, \& Q_3$					
SG	0.024	0.030	0.022	0.035	0.034
TTO	0.040	0.043	0.033	0.048	0.043

Completion time

Completion times were recorded for each session and separately for each health state within each session. Unsurprisingly, for our full sample baseline measurements took longer (5.5 minutes on average) than second individual measurements (little over 3 minutes on average), i.e. repetition decreased time needed for completion ($t(294) = 10.09$, $p < 0.001$). When we focused on subjects in CDM, we observed that group measurements (around 5.5 minutes) took approximately as long as baseline measurement (paired t-test, t(190) = -0.20, p = 0.84). When applying our analytical approach on within-subjects completion times, similar to our analyses on decision outcomes, fixed effects were also obtained for health states separately, to determine if completion times were affected by severity. In turned out that both when estimating the group and carryover effect almost all fixed effects were significant. The only fixed effect that was not significant was that of treatment in the carryover effects model (p=0.28). Collectively, these findings indicated that decision time consistently decreased: from TTO compared to SG, for repeated sessions, for more severe health states. Furthermore,

the group and carryover effect indicated that collective decisions took longer, while subsequent individual measurements were completed faster for subjects in CDM.

Bargaining weights

Finally, we explored the collective decision making process by analyzing decision dynamics within dyads completing the CDM task. We estimated to what extent group QALY weights deviated from QALY weights we observed for the group members at baseline (i.e. I1-CDM). At the aggregate level, a pattern in which the group elicitation falls in-between the two individual estimates is observed most frequently (see Table O.10.C.3). Such a pattern suggests that a majority of groups reached a consensus somewhere in-between their individual estimates (except for TTO-Q1). Nonetheless, outside consensus group utility weights (lower than min, higher than max) are not uncommon and represent between 28 and 43% of the groups, depending on health state and method. When we investigated within-group consensus (i.e. the proportion of consensus across methods and health states), we observed that groups reach consensus in almost two-thirds of elicitations (64.97%). Only two groups (4%) failed to reach consensus on any elicitation on both SG and TTO. We also found no effect of reaching a consensus or not carrying over into subsequent individual decisions in CDM-I2 (t-tests, all p's > 0.18).

Table O.10.C3. Decision process: Location of group utility weight compared to individual weights and median decision weight for high valuators (n = 49).

	SG-Q_1	SG-Q_2	SG-Q_3	TTO-Q_1	TTO-Q_2	TTO-Q_3
Location of utility weight						
Below the min	4	10	11	3	9	8
Above the max	13	10	6	11	12	6
At the min	4	2	2	4	2	4
At the max	7	3	2	15	3	2
In-between	21	24	28	17	23	29
Decision weight	0.89	0.64	0.43	1.00	0.72	0.44

Note: Min and max refer to the lowest and highest individual valuation, respectively.

Next, we estimated the decision weight associated with the higher individual QALY weight in a given group, i.e. the high valuator, for a given decision. We obtained this decision weight by assuming that collective decisions were a weighted summation of individual QALY. In other terms, we calculated decision weight $\propto_H$ of the high valuator in group QALY weights (GQW), by rearranging the following equation: $GQW = \propto_H * IQW_H + (1 - \propto_H) * IQW_L$. Here, IQW_H and IQW_L reflect baseline QALY weights for the high valuator and their partner who assigned lower utility to that health state, respectively. In this context, if $\propto_H > 0.5$ the high valuator has more weight in decisions, while for $\propto_H < 0.5$ the opposite holds. For the sake of clarity, we removed 6 observations corresponding to the cases where the two individuals' QALY weights were identical. Table 3 shows that for the best health state (Q_1), the group tended to follow the individual with the higher utility, whereas the opposite occurred for the worst health state (both for SG and TTO).

References

ABDELLAOUI, M. 2000. Parameter-free elicitation of utility and probability weighting functions. *Management science,* 46**,** 1497-1512.

ABDELLAOUI, M., BLEICHRODT, H., L'HARIDON, O. & VAN DOLDER, D. 2016. Measuring Loss Aversion under Ambiguity: A Method to Make Prospect Theory Completely Observable. *Journal of Risk and Uncertainty,* 52**,** 1-20.

ATTEMA, A. E., BLEICHRODT, H. & L'HARIDON, O. 2018. Measuring ambiguity preferences for health. *Health Economics***,** 1-18.

ATTEMA, A. E. & BROUWER, W. B. 2008. Can we fix it? Yes we can! But what? A new test of procedural invariance in TTO-measurement. *Health Economics,* 17**,** 877-885.

ATTEMA, A. E. & BROUWER, W. B. 2009. The correction of TTO-scores for utility curvature using a risk-free utility elicitation method. *Journal of health economics,* 28**,** 234-243.

ATTEMA, A. E. & BROUWER, W. B. 2010. On the (not so) constant proportional trade-off in TTO. *Quality of Life Research,* 19**,** 489-497.

ATTEMA, A. E. & BROUWER, W. B. 2012a. Constantly proving the opposite? A test of CPTO using a broad time horizon and correcting for discounting. *Quality of Life Research,* 21**,** 25-34.

ATTEMA, A. E. & BROUWER, W. B. 2012b. The way that you do it? An elaborate test of procedural invariance of TTO, using a choice-based design. *The European Journal of Health Economics,* 13**,** 491-500.

ATTEMA, A. E. & BROUWER, W. B. 2014. Deriving time discounting correction factors for TTO tariffs. *Health economics,* 23**,** 410-425.

ATTEMA, A. E., BROUWER, W. B. & L'HARIDON, O. 2013a. Prospect theory in the health domain: a quantitative assessment. *Journal of health economics,* 32**,** 1057-1065.

ATTEMA, A. E., BROUWER, W. B., L'HARIDON, O. & PINTO, J. L. 2016. An elicitation of utility for quality of life under prospect theory. *Journal of health economics,* 48**,** 121-134.

ATTEMA, A. E., VERSTEEGH, M. M., OPPE, M., BROUWER, W. B. & STOLK, E. A. 2013b. Lead time TTO: leading to better health state valuations? *Health economics,* 22**,** 376-392.

BLEICHRODT, H. 2001. Probability weighting in choice under risk: an empirical test. *Journal of Risk and Uncertainty,* 23**,** 185-198.

BLEICHRODT, H., ABELLAN-PERPIÑAN, J. M., PINTO-PRADES, J. L. & MENDEZ-MARTINEZ, I. 2007. Resolving Inconsistencies in Utility Measurement Under Risk: Tests of Generalizations of Expected Utility. *Management Science,* 53**,** 469-482.

BLEICHRODT, H., DOCTOR, J. N., GAO, Y., LI, C., MEEKER, D. & WAKKER, P. P. 2017. Resolving Rabin's Paradox.

BLEICHRODT, H., PINTO, J. L. & WAKKER, P. P. 2001. Making descriptive use of prospect theory to improve the prescriptive use of expected utility. *Management science,* 47**,** 1498-1514.

BLEICHRODT, H., VAN RIJN, J. & JOHANNESSON, M. 1999. Probability weighting and utility curvature in QALY-based decision making. *Journal of mathematical psychology,* 43**,** 238-260.

CULYER, A. J. & WAGSTAFF, A. 1993. Equity and equality in health and health care. *Journal of health economics,* 12**,** 431-457.

DOCTOR, J. N., BLEICHRODT, H. & LIN, H. J. 2010. Health utility bias: a systematic review and meta-analytic evaluation. *Medical Decision Making,* 30**,** 58-67.

FRANCIS, L. J., BROWN, L. B. & PHILIPCHALK, R. 1992. The development of an abbreviated form of the Revised Eysenck Personality Questionnaire (EPQR-A): Its use among students in England, Canada, the USA and Australia. *Personality and individual differences,* 13**,** 443-449.

KAHNEMAN, D. & TVERSKY, A. 1979. Prospect theory: An analysis of decision under risk. *Econometrica***,** 263-291.

KÖBBERLING, V. & WAKKER, P. P. 2005. An index of loss aversion. *Journal of Economic Theory,* 122**,** 119-131.

LAIBSON, D. 1997. Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics,* 112**,** 443-478.

MARTIN, A. J., GLASZIOU, P., SIMES, R. & LUMLEY, T. 2000. A comparison of standard gamble, time trade-off, and adjusted time trade-off scores. *International Journal of Technology Assessment in Health Care,* 16**,** 137-147.

OLIVER, A. 2003. The internal consistency of the standard gamble: tests after adjusting for prospect theory. *Journal of health economics,* 22**,** 659-674.

PERPIÑÁN, J. M. A., MARTÍNEZ, F. I. S., PÉREZ, J. E. M. & MARTÍNEZ, I. M. 2009. Debiasing eq-5d tariffs. New estimations of the Spanish EQ-5D value set under nonexpected utility. Centro de Estudios Andaluces.

PINTO-PRADES, J.-L. & ABELLAN-PERPIÑAN, J.-M. 2012. When normative and descriptive diverge: how to bridge the difference. *Social Choice and Welfare,* 38**,** 569-584.

PLISKIN, J. S., SHEPARD, D. S. & WEINSTEIN, M. C. 1980. Utility functions for life years and health status. *Operations research,* 28**,** 206-224.

RABIN, M. 2000. Risk aversion and expected-utility theory: A calibration theorem. *Econometrica,* 68**,** 1281-1292.

STIGGELBOUT, A. M., KIEBERT, G. M., KIEVIT, J., LEER, J.-W. H., STOTER, G. & DE HAES, J. 1994. Utility assessment in cancer patients: adjustment of time tradeoff scores for the utility of life years and comparison with standard gamble scores. *Medical Decision Making,* 14**,** 82-90.

TANGNEY, J. P., BOONE, A. L. & BAUMEISTER, R. F. 2018. High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Self-Regulation and Self-Control.* Routledge.

TOPLAK, M. E., WEST, R. F. & STANOVICH, K. E. 2011. The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & cognition,* 39**,** 1275.

TVERSKY, A. & KAHNEMAN, D. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of risk and uncertainty,* 5**,** 297-323.

VAN DER POL, M. & ROUX, L. 2005. Time preference bias in time trade-off. *The European Journal of Health Economics,* 6**,** 107-111.

VAN OSCH, S. M., WAKKER, P. P., VAN DEN HOUT, W. B. & STIGGELBOUT, A. M. 2004. Correcting biases in standard gamble and time tradeoff utilities. *Med Decis Making,* 24**,** 511-7.

VERSTEEGH, M. M., VERMEULEN, K. M., EVERS, S. M., DE WIT, G. A., PRENGER, R. & STOLK, E. A. 2016. Dutch tariff for the five-level version of EQ-5D. *Value in health,* 19**,** 343-352.

WAKKER, P. & DENEFFE, D. 1996. Eliciting von Neumann-Morgenstern utilities when probabilities are distorted or unknown. *Management science,* 42**,** 1131-1150.

WAKKER, P. P. 2010. *Prospect theory: For risk and ambiguity*, Cambridge university press.

WALTERS, S. J. & BRAZIER, J. E. 2005. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Quality of life research,* 14**,** 1523-1532.