Gutenberg School of Management and Economics

& Research Unit "Interdisciplinary Public Policy"

Discussion Paper Series

# *Do people choose what makes them happy and how do they decide at all? A theoretical inquiry*

Niklas Scheuer

January 24, 2020

Discussion paper number 2002

Johannes Gutenberg University Mainz
Gutenberg School of Management and Economics
Jakob-Welder-Weg 9
55128 Mainz
Germany
wiwi.uni-mainz.de

Contact Details:

Niklas Scheuer
Chair in Macroeconomics
Johannes Gutenberg University Mainz
Jakob-Welder-Weg 4
55128 Mainz
Germany

scheuer@uni-mainz.de

# Contents

# Do people choose what makes them happy and how do they decide at all? A theoretical inquiry

Niklas Scheuer[(a),1]
[(a)]Department of Economics, Johannes Gutenberg University Mainz

January 24, 2020

We develop a theoretical model that jointly explains optimal choices and happiness. We work with constant elasticity of substitution functions for utility and happiness. Employing a choice framework, individuals are confronted with two options. When there exists a trade-off, we determine parametric conditions for which individual happiness and utility coincide as well as oppose each other. Comparing the empirical evidence of Benjamin et al. (2012), our model can explain three out of four possible happiness-utility combinations. Regarding how individuals actually decide, our findings suggest that this is partly random. This explanation accounts for the remaining 11.2 % of individuals.

Declarations of interest: none

JEL codes: D11, D91, I31
Keywords: Consumer Economics, Theory, General Welfare, Well-Being, Micro-Based Behavioral Economics

## 1 Introduction

Do people choose what makes them happy? Economists observe choices, infer preferences and derive functions reflecting these. These are the well-known utility functions. General economic knowledge tells us that per definition, utility maximization yields optimal behavior. In reality, however, we observe many types of behavior. Hence, some economists follow a less traditional approach. Among these are Benjamin et al. (2012). They examine both principles separately and find that a divergence between optimal choices and choices yielding higher happiness occur.

Hence, how we can rationalize these empirical results using standard economic theory is an open question. While there are many empirical papers on happiness, a theoretical framework to analyze choice-inferred utility and happiness properly appears to be missing.

We develop a model that captures the empirical structure of Benjamin et al. (2012). In their paper, they have different scenarios, each of which is described by a discrete choice set of two options. Every possible option features certain payoffs. Individuals are then required to indicate which option they would choose and with which they would be happier. They showed that, in some cases, a divergence occurs. In order to show this theoretically, we use standard constant elasticity of substitution (CES) functions for happiness and utility, treating both concepts as different from one another. The former is nested in the latter, making the utility function a nested CES function. We only consider economic goods. This entails using strictly positive

[1]Contact details: Niklas Scheuer, Johannes Gutenberg University, Jakob-Welder-Weg 4, 55131 Mainz, Germany, phone +49.6131.39-24701, scheuer@uni-mainz.de. I would like to thank all participants of the Chair Research Days in 2017 and 2019 and of the Brown Bag Seminar in 2019. Special thanks go to Klaus Wälde, Marten Hillebrand, Steffen Altmann and Jean Roch Donsimoni for very specific comments and detailed discussions.

weights inside the happiness and utility function and assuming positive marginal utility. Based on this setup, we determine parametric conditions to show when the evaluation of options in terms of happiness and utility yields either exact matches, supporting the traditional approach, or diverges between the two options, capturing a failure of the traditional approach.

Our contribution to the field of happiness economics is a theoretical framework based on empirical evidence to jointly analyze happiness and utility. Moreover, our model represents a general version, which allows us to analyze more scenarios than the empirical evidence suggests, if model requirements are fulfilled. We further offer explicit conditions to explain the interaction between happiness and utility for various types of bivariate decision problems.

When considering no trade-off scenarios, we find that happiness and utility go into unambiguous directions. Once trade-offs are present, we obtain two trade-off cases, including two subcases each. For *every* potential case, four possible interactions between happiness and utility emerge. For two of those, happiness and utility coincide, which our model can always explain (even though trade-offs are present). However, when they do not match, a CES-framework can only account for one out of two combinations. We provide an extension, where these assumptions are relaxed if we consider economic bads. However, therefrom we abstract as from an empirical standpoint they appear to be irrelevant. A possible explanation for this contradiction is the inability of individuals to correctly evaluate and compare alternatives for various reasons (analysis paralysis, trembling hand, lack of awareness or the like), which leads to deciding randomly. Empirical evidence suggests that this occurs with a share of 11.2 % overall, and ranges from 4 to 32 % for scenarios fitting our assumptions.

We begin with related literature in section 2 and introduce a theoretical framework in section 3, where we also shed more light on the empirical paper. In section 4, we focus on trade-off cases. This is of particular interest for us as it conveys meaningful implications. Section 5 clarifies how individuals actually decide given our framework in combination with empirical findings introduced previously. Section 6 concludes.

# 2 Related literature

First and foremost, happiness and well-being have enjoyed increasing popularity in the field of economics in recent years. Examples of articles published in top journals are Aghion et al. (2016) on creative destruction and its impact on subjective well-being, Adhvaryu et al. (2019) on early life issues and their impact on mental health, Deaton (2018) on the effects of well-being on life-cycle theory and policy, Liberini et al. (2017) on the relationship between voting and happiness, Campante et al. (2015) on religion and its effect on economic growth and happiness or Oswald et al. (2015) on the relationship between happiness and productivity. Literature on happiness and utility can essentially be divided into three categories. We begin with the role of happiness in economics, continue with the historic connection between utility theory and happiness, and, lastly, we consider state-of-the-art theoretical modeling of happiness and decision utility. A brief look at the biological side, which justifies certain economic assumptions, will come last.

Concerning happiness in economics, one of the most influential papers for happiness and economics belongs to Richard Easterlin. He found that despite growing annual income during the seventies in the US, average happiness did not change. This rather unexpected finding marks the starting point of happiness research in economics and is referred to as the Easterlin Paradox (Easterlin, 1974, 1995). Frey and Stutzer (2000, 2002) explicitly differentiate happiness and individual utility. The latter, according to them, only depends on certain factors, derivable from revealed actions and used to explain decisions, whereas happiness is viewed as a broader concept. They also collect data from Germany and Switzerland to support their assumptions and find important determinants of happiness. They identify unemployment and inflation as economic indicators and marital status as a private indicator to be the driving forces of happiness in Germany and Switzerland. Further work on income and happiness comes from Erzo Luttmer (2005), who found that absolute income does not matter for experiencing happiness but the relative position towards your peers does. In other words, he showed that comparisons with neighbors cause happiness if one's own income exceeds that of the neighbor. Kahneman and Deaton (2010) showed that increasing income is associated with higher levels of life evaluation and happiness. However, although life evaluation, which is defined according to Kahneman and Deaton as simply how individuals think about their life, rises steadily with income, happiness does so only until an annual income of $75,000. This indicates that if an individual receives a higher annual income, a further increase in income is not associated with higher levels of happiness, although individuals evaluate their life higher than before.

Concerning happiness and utility theory, the beginning of happiness playing a role in utility theory can be marked as the time utilitarianism was introduced. Back then, Jeremy Bentham, a very prominent supporter, defined utility in various ways. A rather unique one states that utility is the approval or disapproval of every action according to which effect this will have on an individual's happiness (Bentham, 1996). Bentham suggested governments should design policies to promote happiness, such that everyone achieves their highest possible level of happiness. The general consensus about utility and happiness, including supporters such as John Stuart Mill or Daniel Bernoulli, was that utility could be seen as life satisfaction and would be determined by actions causing *pain* or *pleasure*. A rather recent approach from Kahneman, Wakker and Sarin (1997) includes the term "experienced utility," which refers closely to Bentham and the pain or pleasure analogy. It features a comparison between the 'usual' utility function that should predict decisions, inducing the highest utility, and the one comprehending the individual's experiences in terms of utility.

Concerning steps towards harmonizing utility theory and happiness, Rayo and Becker (2007) are among the first to analyze happiness in a theoretical framework. They specify an evolutionary approach in which they use a certain strategy $x$ out of a set of foraging strategies $X$, given a certain state of nature $s$, to obtain an output $y \in \mathbb{R}$. This output is the food you gathered, giving you a certain level of happiness $V(y)$, which is defined according to an innate happiness function $V$ every individual inherits. The expected level of happiness is then defined as $u(x, s)$ and used as a decision criterion for all other strategies. In addition, Rayo und Becker wrote a comment on Stevenson and Wolfers (2008), where they pick a different approach and model utility as depending on happiness $h$ and a commodity $Z$. They assume that both happiness ($\frac{\partial u}{\partial h} > 0$) and the commodity ($\frac{\partial u}{\partial Z} > 0$) increase utility. They also argue that happiness and utility are related, happiness being an argument of the utility function rather than a substitute. Kimball and Willis (2006) defined happiness as consisting of two components, one long-term, which they refer to as *baseline mood*, and one short-term covering instant effects on happiness, which they refer to as *elation*. They view happiness as being part of lifetime utility. Benjamin et al. (2010) follow an approach similar to Rayo and Becker. They differ, however, by assuming that happiness is not a variable itself but depends on many other factors. These will impact utility and happiness directly. Through the latter, they also influence utility indirectly. Hence, they model utility as $u(h(X), X)$. Reflecting the *state of the art* when it comes to happiness and utility theory, we take a different perspective and think of happiness and utility as ultimately two different layers that are inherent in an individual. One where an individual tries to make themself happiest and one where the individual decides in a more rational manner. Every time an individual faces a choice, both layers weigh into the decision, leading to an outcome where happiness and utility either coincide or conflict with each other. This echoes similar frameworks modeling dual selves, such as Fudenberg and Levine (2006) or O'Donoghue and Rabin (1999). The former develop a dual-self model about a short-run and long-run self, that whenever a decision is to be made, determines it as an outcome of both 'selfs' together. The latter described individuals as naive or sophisticated depending on how well they can actually resist temptation and whether they know about their resisting behavior or live in denial.

Concerning happiness and biological factors, only a few economists covered this connection to justify their analysis. Firstly, Rayo and Becker (2007) elaborated on happiness being a bounded measure, since neuro cells can only receive a certain amount of impulses from the nervous system. Hence, it should be modeled accordingly in economics and considered to be bounded. Fliessbach et al. (2007)[2] looked at some brain scans in competition situations with other individuals. They found that as soon as one individual outscores the other, the reward center is activated in our brain. This fits the finding of Luttmer (2005). Weiss et al. (2008) suggest a generic relationship between positive personality traits and happiness traits. Hence, one could argue that genes might play a major role in the variation of people's happiness.

# 3 The model

We begin with providing the single and aggregated evidence on which our model is based, before setting up a framework and conditions for an optimal decision.

## 3.1 The evidence - One scenario

In their original paper, Benjamin et al. (2012) ask study participants (1066 adults composed of 1000 adult Americans and 633 students) about certain hypothetical scenarios, each with two possible outcomes with participants having to choose one of them. Respondents are then asked about their life satisfaction and happiness, as a result of their answer, immediately after they have made their hypothetical choice[3]. The authors then compare responses. One of their main results is that the single best predictor of choice is predicted subjective well-being (as an aggregate of happiness and life satisfaction), but there remain some discrepancies between choice and subjective well-being. These appear to be systematic and meaningful. Put differently, their paper shows empirically that happiness can also be seen as an argument of the utility function instead of a substitute. They show this by comparing certain trade-off scenarios, looking at descriptive statistics and explaining choice using ordinary least squares, probit and ordered probit regressions.

The following table shows the descriptive results for the first scenario that compares income and sleep, which we want to focus on explicitly. There is Option 1 - $80,000 per year salary and 7.5 hours of sleep per day versus more income and less sleep, i.e. Option 2 - $140,000 per year salary and 6 hours of sleep per day. The authors pooled all responses to questions regarding subjective well-being, such as life satisfaction, own happiness or felt happiness together[4], such that the following table is obtained:

| subjective well-being\choice | *Option* 1 | *Option* 2 |
|---|---|---|
| *Option* 1 | 58 % | 12 % |
| *Option* 2 | 1 % | 29 % |

**Table 1** *Data on subjective well-being and choice from Benjamin et al. (2012) for scenario 1 in the Denver group*

[2]We thank Steffen Altmann for making us aware of this literature.

[3]Hence, the subjective well-being measures are only predictions as the choices are hypothetical. This means at the point in time the survey takes place the respondents do not know how the decisions are going to impact their feelings in the future. They can only assume.

[4]There are five subjective well-being questions overall, of which four aim directly at happiness and one targets life satisfaction. For both their samples, subjective well-being is always a mix put together out of these five questions.

A total of 58 % of the survey respondents did not only choose Option 1 (henceforth characterized as $u_1 > u_2$) but are also happier with it (henceforth characterized as $h_1 > h_2$). A total of 29 % of the survey respondents chose Option 2 (henceforth characterized as $u_1 < u_2$) and are also happier with it than with Option 1 (henceforth characterized as $h_1 < h_2$). On the other hand, we can observe that 12 % of respondents would choose Option 2 despite being happier with Option 1. Similarly, 1 % of respondents chose Option 1 despite stating they would be happier with Option 2. Hence, we are looking at a discrete decision problem, where one can select Option 1 or Option 2. This survey evidence for a mismatch between happiness and (assumed 'optimal') choices between income and sleep is further supported by additional scenarios pitting other similar trade-offs against each other. As of now, when we mention 'choices,' we refer to 'optimal choices.'

## 3.2 The evidence - An aggregated view

Benjamin et al. (2012) use two samples in their paper, one is the Cornell National Social Survey and the other comes from patients at a waiting room in Denver. Looking at descriptive results from the Denver study, we can obtain the following:

| subjective well-being\choice | $u_1 > u_2$ | $u_1 < u_2$ |
|---|---|---|
| $h_1 > h_2$ | 41.3 % | 9.3 % |
| $h_1 < h_2$ | 5.7 % | 44.7 % |

**Table 2** *Data on subjective well-being and choice from Benjamin et al. (2012) for all scenarios in the Denver study*

Looking at descriptive results from the Cornell study, we observe a similar pattern of more mass along the main diagonal, as opposed to happiness-choice combinations that do not match:

| subjective well-being\choice | $u_1 > u_2$ | $u_1 < u_2$ |
|---|---|---|
| $h_1 > h_2$ | 44.2 % | 12.1 % |
| $h_1 < h_2$ | 9.3 % | 34.3 % |

**Table 3** *Data on subjective well-being and choice from Benjamin et al. (2012) for all scenarios in the Cornell study*

We see that, most of the time, people's subjective well-being and choices are consistent, as has already been mentioned above. This indicates that our representative scenario captures the relationship between happiness and choice quite well from the beginning, as we see a very similar pattern overall.

We can summarize this in the following table, which provides us with a frame of reference and shows the pattern we wish to explain explicitly. When choice-inferred utility and happiness agree with each other, we use the term 'coincide,' as of now. If they diverge from each other, henceforth, we use the term 'contradict.'

<table>
<tr><th>$h\backslash u$</th><th colspan="2">$u_1 > u_2$</th><th colspan="2">$u_1 < u_2$</th></tr>
<tr><td>$h_1 > h_2$</td><td>coincide</td><td>I</td><td>contradict</td><td>II</td></tr>
<tr><td>$h_1 < h_2$</td><td>contradict</td><td>III</td><td>coincide</td><td>IV</td></tr>
</table>

**Table 4** *Possible combinations of utility and happiness*

This will serve as the foundation to derive our theoretical model[5].

## 3.3 The framework

In this section, we present the case, where utility and happiness are now modeled explicitly. We remain close to the empirical evidence and consider scenarios with two possible options an individual has to choose between and state which he/she prefers in terms of happiness. We later analyze the comparative statics of our framework. This allows us to determine general criteria which characterize an optimal decision for an individual facing two options.

### 3.3.1 Preferences

Considering one scenario, we obtain utility and happiness induced through payoffs from Option 1 $(x_1, y_1)$ and Option 2 $(x_2, y_2)$. We obtain the following general utility function

$$u(h, x, y) = \beta h^\delta + (1 - \beta)\left(\gamma x^\delta + (1 - \gamma)\, y^\delta\right) \text{ with } 0 < \gamma, \beta, \delta < 1, \tag{3.1}$$

where happiness is described by

$$h \equiv h(x, y) = \alpha x^\theta + (1 - \alpha)\, y^\theta \text{ with } 0 < \alpha, \theta < 1. \tag{3.2}$$

We use a standard CES function for happiness with a degree of $\theta$ and a nested CES function for utility. $\delta$ and $\theta$ are constants and inside the unit interval. Furthermore, we identify the following share parameters $\alpha$, $\beta$ and $\gamma$. They are relative weights within the happiness and utility function attached to the option-specific payoffs ($x_1$, $x_2$, $y_1$ and $y_2$). At the outset, we think of $x$ and $y$ as 'economic goods'[6], which is why we exclude negative weights, meaning neither $\alpha$, $\beta$ and $\gamma$ nor $(1 - \alpha)$, $(1 - \beta)$ and $(1 - \gamma)$ can be negative.

Looking at the happiness function, we can see that the larger $\alpha$ is, the more $x$ is emphasized as a driver of happiness. Conversely, for a lower $\alpha$, $y$ affects happiness more strongly. Regarding utility, the larger $\beta$ is, the more happiness is underlined as the key driver of utility, which will lead to a match between happiness and utility. Whereas, if $\beta$ is low, the direct effect of $x$ and $y$ onto utility matters more, relative to happiness from $x$ and $y$. This direct effect can be divided into either $x$ or $y$ being the driver, depending on $\gamma$. From looking at the utility function (3.1), this parameter will influence the direct effect of $x$ and $y$ onto utility. Firstly, if $\gamma$ is large, $x$ is more emphasized, whereas if it is small, $y$ is more strongly emphasized. This is why we are going to analyze several parameter combinations to show under which conditions utility and happiness coincide and when they do not. Given the empirical evidence from above, $\beta$ cannot be equal to 1, as there are individuals who do not choose what makes them happy, i.e. $u \neq h$.

Knowing preferences from above, we are already able to describe happiness and utility where options and their payoffs are without a trade-off. By this, we mean three possible combinations

[5] One might argue that happiness is rather an instant measure, whereas for choice an individual might think more about the long-term consequences, which might lead to the difference that we actually see. This would indicate that you cannot really compare the two due to the different time frames. In order to make this connection clear, one could modify the survey questions and distinguish between a short run and a long run outcome for choice and happiness to assert the importance of the time component which Benjamin et al. did. Unfortunately, due to their pooling together of subjective well-being measures, we cannot clearly identify which measure drives the results we observe. The precise intertemporal relationship between happiness and utility is beyond the scope of this paper, as we choose to focus on static trade-offs that more closely match the evidence available. This approach is in the same spirit as previous work on (behavioral) utility theory, such as Regret Theory (Loomes and Sugden 1982, 1986) or Disappointment Theory (Gul 1991).

[6] At this point, one could think about the elasticity of substitution for utility and happiness. It cannot be determined for utility, as utility is described by a nested CES function. For happiness, the elasticity of substitution reads $\frac{1}{1-\theta}$.

of option-specific payoffs, which will always leave an individual with higher happiness and utility for one particular option. This can be summarized as

**Corollary 1** *There are three cases that do not imply trade-off situations, where strictly positive parameters will always lead to a match between happiness and utility.*
*(i-a) Assuming $x_1$ equals $x_2$, and $y_1$ is larger, respectively, smaller than $y_2$, a strictly positive preference parameter $\alpha$ from (3.2) and $\beta$ from (3.1) imply happiness and utility for Option 1 being larger, respectively, smaller than happiness and utility for Option 2*

$$h_1 > h_2 \wedge u_1 > u_2 \Leftrightarrow 0 < \alpha, \beta < 1 \vee$$
$$h_1 < h_2 \wedge u_1 < u_2 \Leftrightarrow 0 < \alpha, \beta < 1.$$

*(i-b) Assuming $y_1$ equals $y_2$, and $x_1$ is larger, respectively, smaller than $x_2$, a strictly positive preference parameter $\alpha$ from (3.2) and $\beta$ from (3.1) imply happiness and utility for Option 1 being larger, respectively, smaller than happiness and utility for Option 2*

$$h_1 > h_2 \wedge u_1 > u_2 \Leftrightarrow 0 < \alpha, \beta < 1 \vee$$
$$h_1 < h_2 \wedge u_1 < u_2 \Leftrightarrow 0 < \alpha, \beta < 1.$$

*(ii) If both option-specific payoffs go into the same direction, i.e. $x_1$ is larger, respectively smaller than $x_2$ and $y_1$ is larger, respectively smaller than $y_2$, then*

$$h_1 > h_2 \wedge u_1 > u_2 \Leftrightarrow 0 < \alpha, \beta < 1 \vee$$
$$h_1 < h_2 \wedge u_1 < u_2 \Leftrightarrow 0 < \alpha, \beta < 1$$

*i.e. the happiness and utility of Option 1 are larger, respectively, smaller than happiness and utility of Option 2.*

Later, we also describe what happens if we relax previously made assumptions regarding the weighting parameters, showing the plurality of our results.

### 3.3.2 Optimal decision

This section discusses the optimal decision regarding happiness and utility. Both are considered to be ordinal measures[7].

At first, we determine the general condition under which an individual is happier with Option 1 compared to 2. This reads

$$h_1 > h_2. \tag{3.3}$$

If we insert the respective functions, rearrange regarding $\alpha$ and consider implications from different values of $x$ and $y$ for different options and scenarios, using the condition above in conjunction with the happiness function in (3.2), we obtain,

$$h_1 > h_2 \Leftrightarrow \alpha > \frac{\Theta_y}{\Theta_y - \Theta_x} \wedge \Theta_y - \Theta_x > 0 \quad \vee \tag{3.4a}$$
$$\Leftrightarrow \alpha < \frac{\Theta_y}{\Theta_y - \Theta_x} \wedge \Theta_y - \Theta_x < 0, \tag{3.4b}$$

[7]We evaluate happiness and utility for both options and determine which option yields more happiness and which yields more utility. Hence, a cardinal measure is irrelevant as the comparison matters. This is why happiness and utility are both ordinal measures.

where the last line employed

$$\Theta_y \equiv y_2^\theta - y_1^\theta \wedge \Theta_x \equiv x_2^\theta - x_1^\theta. \tag{3.5}$$

This condition is a general result, as it depends on the relationship between $x$ and $y$ payoffs.

We next consider the analogous parameter condition for the utility function, in order to observe utility from 1 being larger than that from 2. Thus, we observe conditions under which an individual chooses Option 1 over 2. This condition is given by

$$u_1 > u_2. \tag{3.6}$$

If we replace the utility expressions with the corresponding equations and assumptions made above, rearrange regarding $\beta$ and consider different value combinations of $x$ and $y$ for each scenario, we can write

$$u_1 > u_2 \Leftrightarrow \beta > \frac{\gamma\Delta_x + (1-\gamma)\Delta_y}{\Delta_h(\alpha) + \gamma\Delta_x + (1-\gamma)\Delta_y} \wedge \Delta_h(\alpha) > -\left[\gamma\Delta_x + (1-\gamma)\Delta_y\right] \quad \vee \tag{3.7a}$$

$$\Leftrightarrow \beta < \frac{\gamma\Delta_x + (1-\gamma)\Delta_y}{\Delta_h(\alpha) + \gamma\Delta_x + (1-\gamma)\Delta_y} \wedge \Delta_h(\alpha) < -\left[\gamma\Delta_x + (1-\gamma)\Delta_y\right], \tag{3.7b}$$

where the last line employed

$$\Delta_y \equiv y_2^\delta - y_1^\delta, \Delta_x \equiv x_2^\delta - x_1^\delta \wedge \Delta_h(\alpha) \equiv h_1(\alpha)^\delta - h_2(\alpha)^\delta. \tag{3.8}$$

Now, (3.7a) and (3.7b) are general parameter restrictions. However, at this stage, we have not yet considered explicit parameter values which would fix the sign of the denominator. This is why we will make distinctions leading to various cases and circumstances under which individuals choose Option 1 over 2 or vice versa.

Concerning the numerator, we can also derive parameter restrictions for $\gamma$, which will help us in our analysis. For certain values of $\gamma$, the difference between differences $\Delta_x$ and $\Delta_y$ will be positive, and for some values of $\gamma$, it will be negative. Hence, depending on $\gamma$, the numerator of the fraction in (3.7a) and (3.7b) is positive or negative leading to two distinct cases. This also becomes essential concerning the ability to explain scenarios using this framework. We establish from the numerator of (3.7a) and (3.7b)

$$\gamma\Delta_x + (1-\gamma)\Delta_y \gtreqless 0 \Leftrightarrow \gamma \gtreqless \frac{-\Delta_y}{\Delta_x - \Delta_y}. \tag{3.9}$$

This allows us to show for which $\gamma$ the comparison of direct effect differences ($\Delta_x$ and $\Delta_y$) is positive or negative. Whether it matters or not depends on $\beta$ and its value.

As we know the preferences and the general conditions for an optimal decision, we can look at happiness and utility in greater detail. Firstly, we cover cases where the options and their payoffs impose a trade-off. By this, we mean one payoff in an option will be larger than its counterpart. Thus, one payoff will be larger in Option 1 than in Option 2 and one payoff will be larger in Option 2 than in Option 1. We then observe happiness and utility, allowing us to describe individuals who choose an option they are also happiest with and those who do not.

# 4 Trade-off

Individuals now face payoffs in such a way that one payoff in Option 1 is larger than its counterpart in Option 2 and one payoff in Option 2 is larger than its counterpart in Option 1. We generally have *two* possible trade-off cases, namely

$$\text{trade-off case 1: } x_1 < x_2 \wedge y_1 > y_2 \text{ or} \tag{4.1}$$

$$\text{trade-off case 2: } x_1 > x_2 \wedge y_1 < y_2. \tag{4.2}$$

Given this structure, many subcases emerge. We begin by analyzing happiness before we continue with utility. The analysis for utility is very comprehensive, which is why, we focus there on trade-off case 1 explicitly, as the second trade-off case can be analyzed analogously. We then illustrate how individuals actually decide and show the explanatory power of our framework in the next section.

## 4.1 Happiness

In the presence of a trade-off between options, we can obtain a threshold value for $\alpha$, denoted $\alpha^*$, which will determine under which conditions the individual is happier. This condition is summarized in lemma 1 below.

**Lemma 1** *When there is a trade-off between Option 1 and Option 2, a unique value $\alpha^*$ exists for which the individual is indifferent between two options in terms of happiness. That value follows from equation (3.4a) and (3.4b), and reads*

$$\alpha^* \equiv \frac{\Theta_y}{\Theta_y - \Theta_x}, \text{ where } \Theta_y \equiv y_2^\theta - y_1^\theta \wedge \Theta_x \equiv x_2^\theta - x_1^\theta, \tag{4.3}$$

*and will always be between* 0 *and* 1*.*

**Proof.** See appendix. ■

This allows us to draw the following proposition:

**Proposition 1** *Cases including trade-off situations are described by option-related payoffs going in opposite directions.*
*(i) We get*

$$h_1 > h_2 \Leftrightarrow \alpha < \alpha^*,$$

*for the trade-off* (4.1) *mentioned above, i.e. the happiness of Option 1 will be larger, respectively, smaller than the happiness of Option 2 if and only if the preference parameter $\alpha$ is smaller, respectively, larger than the cutoff value $\alpha^*$.*
*(ii) We get*

$$h_1 > h_2 \Leftrightarrow \alpha > \alpha^*$$

*for the trade-off* (4.2) *mentioned above. The happiness of Option 1 will be larger, respectively, smaller than the happiness of Option 2 if and only if the preference parameter $\alpha$ is larger, respectively smaller than the cutoff value $\alpha^*$.*

**Proof.** See appendix. ■

For trade-off case (4.1), we know that the value of $y$ for Option 1 promises a higher payoff than Option 2, whereas for the value of $x$, it is just the opposite. In order to be happier with Option 1, we need to emphasize the value of $y$ sufficiently.[8] Thus, Option 1 leads to a higher level of happiness in case $\alpha$ is smaller than $\alpha^*$ such that $x_2$, which exceeds $x_1$, does not impact the happiness function as much as $y_1$, which exceeds $y_2$.

On the other hand, if we are considering trade-off case (4.2), $x_1$ exceeds $x_2$ and $y_2$ is larger than $y_1$. Hence, $\alpha$ needs to be larger than $\alpha^*$ in order for $x_1$ to be emphasized more strongly than $y_2$ such that Option 1 creates a higher level of happiness.

[8]For a better understanding keep in mind that our happiness equation reads

$$h = \alpha x^\theta + (1 - \alpha)\, y^\theta.$$

## 4.2 Utility

Concerning our utility function, we determine $\beta$ for which an individual either prefers Option 1 or Option 2. As we consider cases that include a trade-off, it is important to discuss the cutoff value for utility beforehand, i.e. $\beta^*$ and the influence of $\gamma$. Later, we explore two subcases of the first trade-off case.

### 4.2.1 Properties of the cutoff value

This section deals with $\beta^*$ and its derivation and impacts. It analyzes $\beta$ where individuals are indifferent, shows the importance of the numerator and denominator of (3.7a) or (3.7b) and covers the impact $\alpha$ has in that regard.

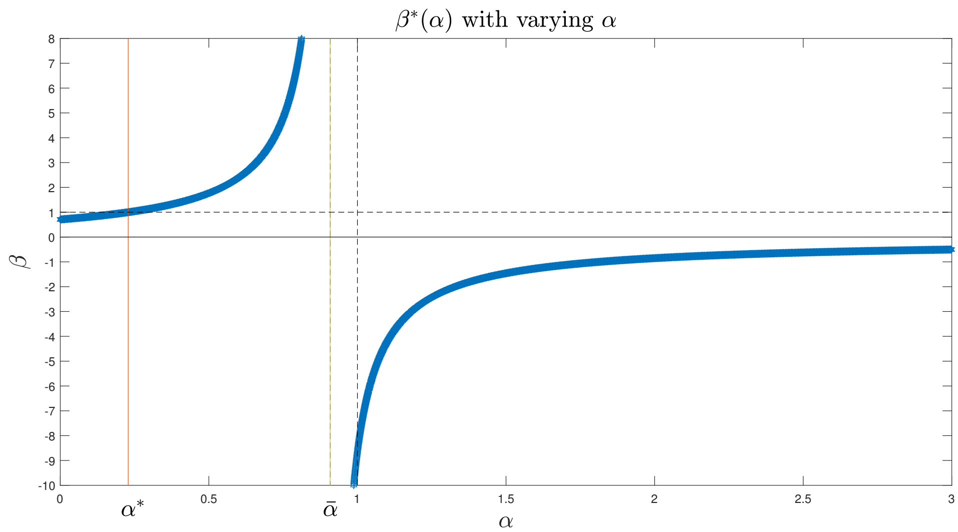
Concerning choice indifference, we can draw the following lemma:

**Lemma 2** *In the case where trade-offs are present between Option 1 and Option 2, a value $\beta^*$ exists for which the individual is indifferent between two options in terms of utility. That value follows from (3.7a), respectively (3.7b), and results from*

$$u_1 = u_2 \Leftrightarrow \beta = \beta^* (\alpha) \text{ with}$$
$$\beta^* (\alpha) \equiv \frac{\gamma \Delta_x + (1-\gamma) \Delta_y}{\Delta_h (\alpha) + \gamma \Delta_x + (1-\gamma) \Delta_y} \equiv \frac{\gamma \Delta_x + (1-\gamma) \Delta_y}{f(\alpha)}. \tag{4.4}$$

**Proof.** See appendix. ■

Apart from depending on the respective differences and $\gamma$, it varies with $\alpha$ as this will have implications on $\Delta_h (\alpha)$ and, eventually, on $\beta^*$. In order to get a feeling for the cutoff value and its dependence on $\alpha$, we depict $\beta^* (\alpha)$ for the first subcase of trade-off case (4.1) with *reasonable values*[9] of $x$ and $y$. For the latter, we took the option-specific payoffs of Benjamin et al. (2012) above and monotonically transformed them appropriately regarding their time horizon. We get




[9]From this point on, if we refer to *reasonable scaling* or *reasonable values*, both values $x$ and $y$ are taken from a scenario of Benjamin et al. (2012) and modified such that the measurement of the original option-specific payoffs is more realistic in terms of units of measurement. Consider, for instance, a scenario with two options in which each contain a monetary payoff measured in monetary units per year, and a 'time' payoff measured in hours. It would be more appropriate to compare over the same time horizon, i.e. to break down the monetary payoff into units per *hour* for both options. In this way, we ensure that a comparison is even more meaningful than before and we do not lose any information or frame any outcome, as it is a simple monotonic transformation. One could also use the raw values from their paper, as they are reasonable but do not imply a good foundation on which an individual can compare options reasonably with each other. This unfortunately holds for almost every scenario suggested by Benjamin et al. (2012), which is why monotonic transformations appear helpful.

**Figure 1** *$\beta^*$ given a subcase of (4.1) and reasonable levels of $x$ and $y$*

The value of $\alpha$ where a pole occurs, i.e. $\bar{\alpha}$, will be explained in detail below.

Focusing on the numerator of $\beta^*(\alpha)$ first, all differences within indicated by $\Delta$ emerge as either positive or negative. Given, for instance, trade-off (4.1) and using (3.8), we see that $\Delta_y$, which was defined as the difference between $y_2^\delta$ and $y_1^\delta$, will turn out negative. $\Delta_x$, i.e. the difference between $x_2^\delta$ and $x_1^\delta$, will turn out positive. If we just compare these, $\Delta_y$ is always smaller than $\Delta_x$, as the former is negative and the latter is positive. So in order to decide whether the numerator of $\beta^*(\alpha)$ will be positive or negative, which influences our analysis greatly, we can compare absolute values to show which difference dominates.

This is where $\gamma$ becomes meaningful, as a different weight is attached to $\Delta_x$ and $\Delta_y$ for different values of $\gamma$. Hence, we distinguish *two* subcases, which are characterized by the numerator being larger (subcase a) or smaller (subcase b) than 0. This goes back to (3.9) and illustrates a $\gamma$ between 0 and 1 for a trade-off case. Considering our assumptions and conditions for $\gamma$, we can say that for a positive numerator, we need

$$\begin{aligned} \gamma\Delta_x + (1-\gamma)\Delta_y &> 0 \\ \gamma &> \frac{-\Delta_y}{\Delta_x - \Delta_y}, \end{aligned}$$

leading to an overall positive direct effect of the differences of $x$ and $y$. Subsequently, the numerator is negative when

$$\begin{aligned} \gamma\Delta_x + (1-\gamma)\Delta_y &< 0 \\ \gamma &< \frac{-\Delta_y}{\Delta_x - \Delta_y}. \end{aligned}$$

This is necessary because $x$ and $y$ are not going to change throughout the analysis once an individual faces a specific trade-off, but preference parameters such as $\alpha$ and $\gamma$ take on different values if we consider different individuals. Hence, this changes $\beta^*$ strongly across individuals, implying that our analysis can be viewed, to some extent, as an analysis of the distribution of parameters.

Concerning the denominator of $\beta^*(\alpha)$, it is inevitable $\alpha$ and its impact are described in greater detail as it plays a crucial role. Differences within trade-off cases oppose each other, which not only affects the numerator (as mentioned above). The denominator of $\beta^*(\alpha)$ was introduced in (4.4), which is a function of $\alpha$ for given values of $x$ and $y$ and reads

$$f(\alpha) \equiv \Delta_h(\alpha) + \gamma\Delta_x + (1-\gamma)\Delta_y, \tag{4.5}$$

where the differences are again taken from (3.8). The denominator is positive for a given $\alpha$ such that the happiness difference exceeds or complements the direct effect of $x$ and $y$ depending on its sign. This direct effect depends on given differences and assumptions regarding $\gamma$. However, for now, we concentrate on $\alpha$, as $\gamma$, once chosen, remains constant and will only be used for making distinctions between subcases. We can generally write

$$f(\alpha) \gtrless 0 \Leftrightarrow \Delta_h(\alpha) \gtrless -[\gamma\Delta_x + (1-\gamma)\Delta_y]. \tag{4.6}$$

Given the structure of the denominator, an $\alpha$ also exists for which the denominator equals 0. Hence, we obtain an $\bar{\alpha}$, for which $\Delta_h(\bar{\alpha})$ will exactly offset the other two differences leading the denominator of $\beta^*(\alpha)$ to be 0.

Concerning $\bar{\alpha}$, that has already appeared in the figure for $\beta^*(\alpha)$ above, we look at (4.5) first. The only factor left impacting the denominator for given differences of $x$ and $y$ is $\Delta_h(\alpha)$.

If that varies, the denominator will equal 0 at some point, leading $\beta^*$ to tend towards $\pm\infty$. This depends strongly on the sign of the numerator and describes the pole mentioned above. However, $\bar{\alpha}$ does not necessarily appear between 0 and 1, as it obviously depends on the values of the differences $\Delta_x$ and $\Delta_y$. Quantitatively speaking, if we assume values for a more reasonable comparison, $\bar{\alpha}$ often occurs within 0 and 1 for various scenarios. If we do not, $\bar{\alpha}$ lies outside of the unit interval. Attempting to determine exactly $\bar{\alpha}$, we write

$$f(\bar{\alpha}) = 0 \Leftrightarrow \Delta_h(\alpha) + \gamma\Delta_x + (1-\gamma)\Delta_y = 0. \tag{4.7}$$

Inserting the happiness difference defined above yields

$$f(\bar{\alpha}) = 0 \Leftrightarrow \left(\bar{\alpha}x_1^\theta + (1-\bar{\alpha})\,y_1^\theta\right)^\delta - \left(\bar{\alpha}x_2^\theta + (1-\bar{\alpha})\,y_2^\theta\right)^\delta + \gamma\Delta_x + (1-\gamma)\Delta_y = 0.$$

Given our structure solving analytically is not possible, but can only be done numerically from this point on. Unfortunately, this is not trivial. If we determine the numerical solution for a given combination of $x$ and $y$ reflecting a certain trade-off case, we obtain results which not only alter with a varying $\alpha$, but are highly dependent on $x$ and $y$. This becomes evident as we calculate the numerical solution with payoffs suggested by Benjamin et al. (2012) for their income and sleep example. Thereby we vary income by breaking down yearly into monthly, daily and hourly income. Using these variations on the time scale used for measuring income, it remains difficult to obtain meaningful solutions. Regarding daily, monthly or yearly income, complex numbers arise for certain values of $\alpha$, which are generally unwanted. This is because a larger value of $x$ or $y$ can lead to a negative $h$ for which no real root can be extracted leading to imaginary numbers.

In order to get an impression of the direction the denominator takes with a varying $\alpha$, we consider the derivative of $f(\alpha)$. Based on previous analysis we expect a negative derivative for trade-off case 1. We know $f(\alpha)$ reads:

$$f(\alpha) = \left(\alpha x_1^\theta + (1-\alpha)\,y_1^\theta\right)^\delta - \left(\alpha x_2^\theta + (1-\alpha)\,y_2^\theta\right)^\delta + \gamma\Delta_x + (1-\gamma)\Delta_y.$$

Given that we derive regarding $\alpha$, the change of the denominator is entirely described through the change of the happiness difference. We obtain

$$\begin{aligned}\frac{d}{d\alpha}f(\alpha) = f'(\alpha) &= \delta\left(x_1^\theta - y_1^\theta\right)\left(\alpha x_1^\theta + (1-\alpha)\,y_1^\theta\right)^{\delta-1} - \delta\left(x_2^\theta - y_2^\theta\right)\left(\alpha x_2^\theta + (1-\alpha)\,y_2^\theta\right)^{\delta-1} \\ &= \delta\left[\left(x_1^\theta - y_1^\theta\right)\frac{1}{h_1(\alpha)^{1-\delta}} - \left(x_2^\theta - y_2^\theta\right)\frac{1}{h_2(\alpha)^{1-\delta}}\right].\end{aligned} \tag{4.8}$$

As before, we cannot tell the sign of the derivative by just looking at the expression. This is because we know the general relationship between $x_1$ ($y_1$) and $x_2$ ($y_2$) for every trade-off *but not* between $x_1$ ($x_2$) and $y_1$ ($y_2$), which would be necessary to determine whether $(x_1^\theta - y_1^\theta)$ or $(x_2^\theta - y_2^\theta)$ are positive or negative. If we consider our sleep and income example, this makes sense as an exact comparison between daily income and the hours of sleep per day appears slightly abstract.

Before analyzing cases, we have to determine the order in which $\alpha^*$ and $\bar{\alpha}$ occur. This will change $\Delta_h(\alpha)$, $f(\alpha)$ and $\beta^*(\alpha)$ accordingly. We obtain the following lemma:

**Lemma 3** *Given trade-off cases (4.1) or (4.2) we can distinguish between subcase a) and b) as the difference between both is* $\gamma\Delta_x + (1-\gamma)\Delta_y$ *larger, respectively smaller than* 0*. If* $\gamma\Delta_x + (1-\gamma)\Delta_y > 0$*, we can identify the following relationship* $\alpha^* < \bar{\alpha}$ *(for (4.2) subcase a):* $\bar{\alpha} < \alpha^*$*). If* $\gamma\Delta_x + (1-\gamma)\Delta_y < 0$ *we obtain* $\bar{\alpha} < \alpha^*$ *(for (4.2) subcase b):* $\alpha^* < \bar{\alpha}$*).*

**Proof.** See appendix. ■

Now we can specifically differentiate between the *two* subcases mentioned earlier.

### 4.2.2 Trade-off (4.1) - Subcase a)

The first subcase, denoted a) from (4.1), characterized by the following conditions, is

$$x_1 < x_2 \wedge y_1 > y_2 \wedge \gamma > \frac{-\Delta_y}{\Delta_x - \Delta_y}. \tag{4.9}$$

This is exactly what we assumed for trade-off (4.1) combined with $\gamma$ taking on a value such that $\gamma$ is above its threshold value. We, furthermore, want to show when $\Delta_h(\alpha)$, then $f(\alpha)$ and, ultimately, $\beta^*(\alpha)$ change signs. Applying the conditions set in (4.9) to proposition 2 and lemma 3, we are able to determine conditions for the values of $\alpha$ introduced earlier, i.e. $\alpha^*$ and $\bar{\alpha}$:

$$\begin{aligned} \Delta_h(\alpha) > 0 \Leftrightarrow \alpha < \alpha^* \; ; \; f(\alpha) \wedge \beta^*(\alpha) > 0 \Leftrightarrow \alpha < \bar{\alpha} \\ \Delta_h(\alpha) < 0 \Leftrightarrow \alpha > \alpha^* \; ; \; f(\alpha) \wedge \beta^*(\alpha) < 0 \Leftrightarrow \alpha > \bar{\alpha} \end{aligned} \tag{4.10}$$

The left-hand side of (4.10) is known from the happiness section of trade-off (4.1). It depicts $h_1 > h_2$ for the upper and $h_1 < h_2$ for the lower line. The right-hand side of (4.10) shows the sign of the denominator, i.e. $f(\alpha)$, and the cutoff, i.e. $\beta^*(\alpha)$, for a given $\alpha$. The impact $\alpha$ has on the denominator can also be depicted as:

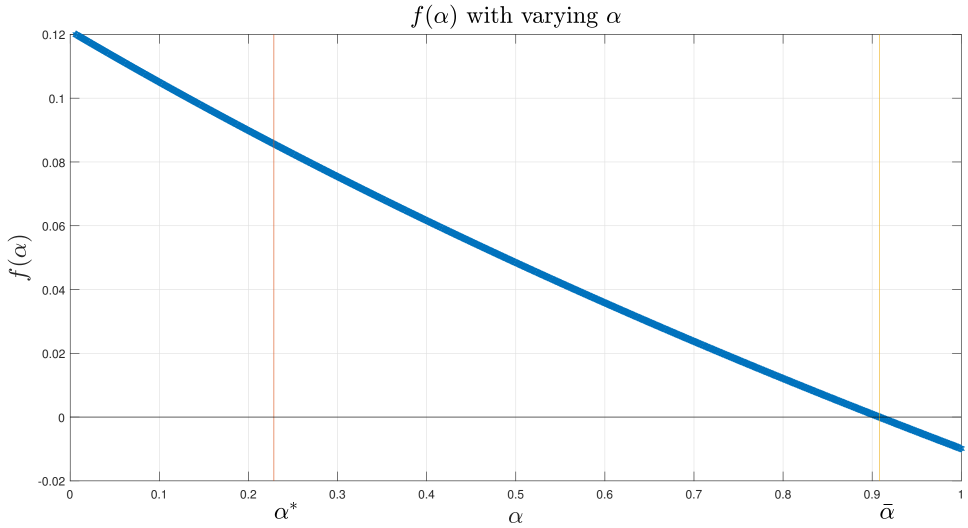



**Figure 2** *$f(\alpha)$ given conditions in (4.9)*

As $\alpha$ passes $\alpha^*$, this has an influence on the value of $f(\alpha)$ and $\beta^*(\alpha)$ but will not change the algebraic sign of the denominator nor of the cutoff. This happens at $\bar{\alpha}$ which can be identified according to the figure as 0.91 for representative values.

We know from lemma 3 that for trade-off (4.1) subcase a) the relationship $\alpha^* < \bar{\alpha}$ has to hold. If we let $\alpha$ increase from 0 upwards, we can distinguish three sets of $\alpha$-values having different impacts on $\Delta_h$, $\beta^*(\alpha)$ and, ultimately, on the utility level between options $u_1 \gtrless u_2$. If we were not adapting the values, we would only face two sets of $\alpha$-values, as $\bar{\alpha}$ would be outside of 0 and 1, which does not change the analysis *per se* but only the number of $\alpha$-sets we consider. The rationale is that for every area and its corresponding boundaries we choose a value of $\alpha$ within, enabling us to analyze its implications. Hence, we define the following relationships:

**Definition 1** *Under the assumptions of (4.9), proposition 1, lemmata 1,2 and 3, areas con-*

*taining strictly positive values of $\alpha$ and $\beta$ are defined as follows*

| *$\alpha$-interval* | *$\beta$-interval* | *Area* |
|---|---|---|
| $\alpha \in (0, \alpha^*)$ | $\beta \in (0, \beta^*)$ | $\mathbb{A}$ |
| | $\beta \in (\beta^*, 1)$ | $\mathbb{B}$ |
| $\alpha \in (\alpha^*, \bar{\alpha})$ | $\beta \in (0, 1)$ | $\mathbb{C}_1$ |
| $\alpha \in (\bar{\alpha}, 1)$ | | $\mathbb{C}_2$ |

If $\bar{\alpha}$ were not to appear between 0 and 1, then $\mathbb{C} = \mathbb{C}_1 \cup \mathbb{C}_2$ holds for the interval where $\alpha \in (\alpha^*, 1)$. Now, we can find a different $\Delta_h$, $f(\alpha)$ and $\beta^*(\alpha)$ for every area. This is why the conditions for choice change within every area and, therefore, have to be considered accordingly. If we put this all together, we obtain

| Option 1 chosen ($u_1 > u_2$) | Option 2 chosen ($u_1 < u_2$) | Value of $\beta^*(\alpha)$ |
|---|---|---|
| $\beta > \beta^*(\alpha)$ ; $\mathbb{A}$ | $\beta < \beta^*(\alpha)$ ; $\mathbb{B}$ | $0 < \beta^*(\alpha) < 1$ |
| $\beta > \beta^*(\alpha)$ | $\beta < \beta^*(\alpha)$ ; $\mathbb{C}_1 \vee \mathbb{C}$ | $1 < \beta^*(\alpha)$ |
| $\beta < \beta^*(\alpha)$ | $\beta > \beta^*(\alpha)$ ; $\mathbb{C}_2$ | $\beta^*(\alpha) < 0$ |

**Table 5** *Overview of areas, choices and values of $\beta^*(\alpha)$ for conditions from (4.9)*

Visualizing the last table, we obtain

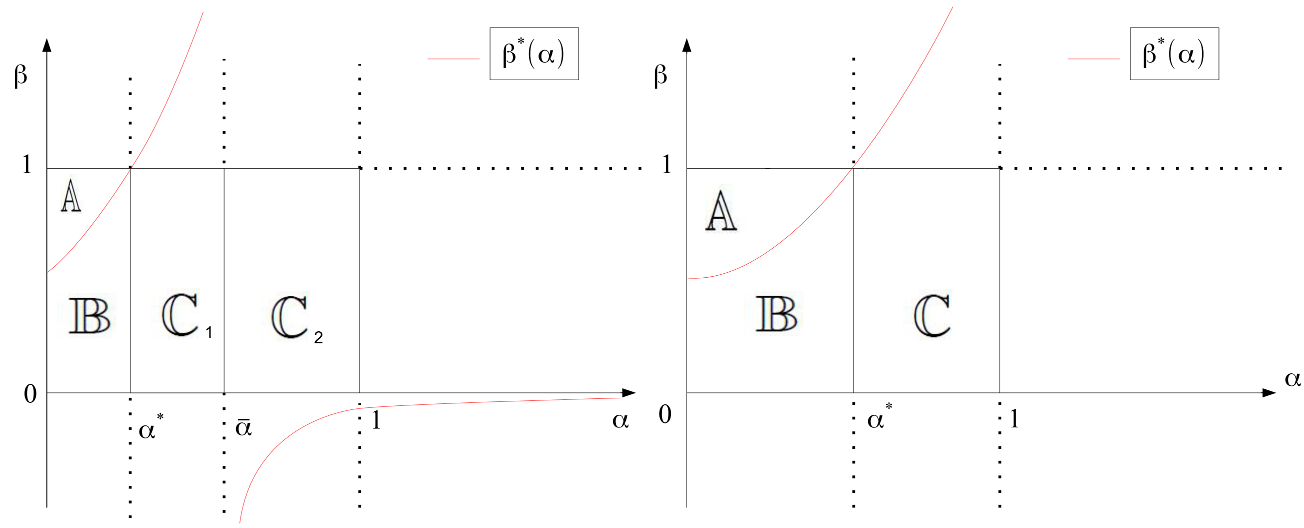



**Figure 3** *Sketch of $\beta^*$ for case (4.1) subcase a) with $\bar{\alpha} \in (0, 1)$ (left-hand side) and with $\bar{\alpha} \notin (0, 1)$ (right-hand side)*

If $\alpha$ is smaller than $\alpha^*$ and $\bar{\alpha}$, we obtain a $\beta^*(\alpha)$ which is between 0 and 1. This allows us to differentiate clearly between $\beta > \beta^*(\alpha)$, i.e. $\mathbb{A}$, for which Option 1 is chosen, and $\beta < \beta^*(\alpha)$, i.e. $\mathbb{B}$, for which Option 2 is chosen. Put differently, if $\alpha$ is reasonably small such that Option 1 is preferred in terms of happiness, $\beta$ needs to be sufficiently high to put enough weight on happiness such that its influence will cause individuals to choose Option 1. If $\beta$ is not high enough, i.e. $\beta < \beta^*(\alpha)$, the weight of $x$ and $y$ inside $u$ is larger than for happiness. This will lead to preferring Option 2 over 1, although individuals would be happier with Option 1.

On the other hand, if an individual is happier with Option 2, i.e. $\alpha > \alpha^*$, there is no $\beta$ between 0 and 1 such that an individual would choose Option 1 over Option 2. In other words, if $\alpha$ passes $\alpha^*$ we enter area $\mathbb{C}_1$ and $\beta^*(\alpha)$ is above 1. This indicates that a $\beta$ larger than $\beta^*(\alpha)$ is a possibility for which we do not allow. Hence, for our conditions above and a strictly positive $\beta$, Option 2 will always be preferred. Similarly for area $\mathbb{C}_2$, $\beta$ cannot be smaller than a negative $\beta^* (\alpha)$, assuming strictly positive parameters. Thus, Option 2 will always be preferred for a $\beta$ between 0 and 1.

In a case where $\bar{\alpha}$ is not inside the unit interval, the differences $\Delta_x$ and $\Delta_y$ are usually rather 'far' apart. There are then three areas of $\alpha$ and $\beta$ values, i.e. $\mathbb{A}$ and $\mathbb{B}$ or $\mathbb{C}$. If $\alpha$ is smaller than $\alpha^*$, $\beta$ is either in $\mathbb{A}$ or $\mathbb{B}$. If $\alpha$ is larger than $\alpha^*$, $\beta$ is part of area $\mathbb{C}$ (see right-hand side of Figure 3).

**Proposition 2** *Given definition 1 and that conditions (4.9) for trade-off (4.1) subcase a) hold, if weights $\alpha$ and $\beta$ take on values such that we are in area*

$$\begin{bmatrix} \mathbb{A}, \\ \mathbb{B}, \\ \mathbb{C} \vee \mathbb{C}_1 \wedge \mathbb{C}_2, \end{bmatrix} \textit{utility from Option 1 is always} \begin{bmatrix} \textit{larger} \\ \textit{smaller} \\ \textit{smaller} \end{bmatrix} \textit{than utility from Option 2.}$$

**Proof.** See appendix. ■

### 4.2.3 Trade-off (4.1) - Subcase b)

Everything referring to utility analysis is going to change for the case where $x_1 < x_2$ and $y_1 > y_2$ with the numerator being negative, as only the interpretation regarding $\alpha$ and $\alpha^*$ stays as the trade-off itself is the same. The comparison of differences regarding the direct effect of $x$ and $y$ is now negative, i.e. $\gamma$ changed. This can be identified by the rightmost expression of

$$x_1 < x_2 \wedge y_1 > y_2 \wedge \gamma < \frac{-\Delta_y}{\Delta_x - \Delta_y}. \tag{4.11}$$

Applying proposition 2 and lemma 3, we describe when $\Delta_h (\alpha)$, $f (\alpha)$ and $\beta^* (\alpha)$ change signs for $\alpha^*$ and $\bar{\alpha}$:

$$\begin{array}{lcl} \Delta_h (\alpha) > 0 \Leftrightarrow \alpha < \alpha^* & ; & f (\alpha) > 0 \wedge \beta^* (\alpha) < 0 \Leftrightarrow \alpha < \bar{\alpha} \\ \Delta_h (\alpha) < 0 \Leftrightarrow \alpha > \alpha^* & ; & f (\alpha) < 0 \wedge \beta^* (\alpha) > 0 \Leftrightarrow \alpha > \bar{\alpha} \end{array}$$

The left-hand side is known from the happiness section of this particular case. It depicts $h_1 > h_2$ for the upper and $h_1 < h_2$ for the lower line. The right-hand side shows the sign of the denominator, i.e. $f (\alpha)$, and the cutoff, i.e. $\beta^* (\alpha)$, for a certain $\alpha$. The impact $\alpha$ has on the denominator can also be depicted as

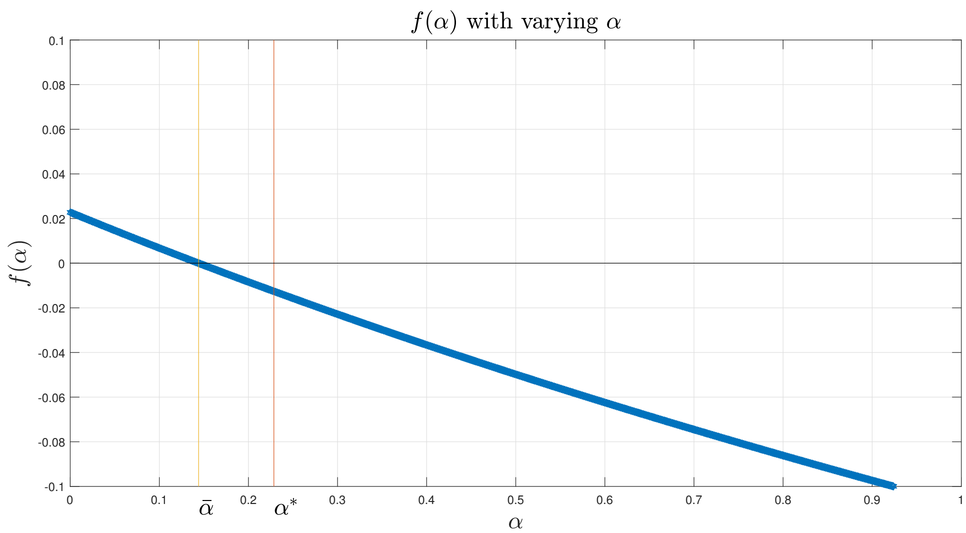



**Figure 4** *$f (\alpha)$ given conditions in (4.11)*

As $\alpha$ passes $\bar{\alpha}$, it affects $f(\alpha)$ and $\beta^*(\alpha)$ and will change their sign. Given the picture from above, $\bar{\alpha} = 0.14$ assuming representative values.

For case (4.1) subcase b) and reasonable scaling we can identify the following relationship $0 < \bar{\alpha} < \alpha^* < 1$. If we let $\alpha$ increase from 0, we can distinguish three sets of $\alpha$ values that impact $\Delta_h$, $\beta^*$ and, ultimately, the utility level between options $u_1 \gtrless u_2$ differently. For every set we choose a value of $\alpha$ within the corresponding boundaries to analyze implications on the other variables. We define the following relationships:

**Definition 2** *Under the assumptions of (4.11), proposition 1, lemmata 1, 2 and 3, areas containing strictly positive values of $\alpha$ and $\beta$ are defined as follows*

| *$\alpha$-interval* | *$\beta$-interval* | *Area* |
|---|---|---|
| $\alpha \in (0, \bar{\alpha})$ | $\beta \in (0,1)$ | $\mathbb{A}_1$ |
| $\alpha \in (\bar{\alpha}, \alpha^*)$ | | $\mathbb{A}_2$ |
| $\alpha \in (\alpha^*, 1)$ | $\beta \in (0, \beta^*)$ | $\mathbb{B}$ |
| | $\beta \in (\beta^*, 1)$ | $\mathbb{C}$ |

If $\bar{\alpha}$ were not to appear between 0 and 1, then $\mathbb{A} = \mathbb{A}_1 \cup \mathbb{A}_2$ holds for the $\alpha$-interval of $\alpha \in (0, \alpha^*)$. We can find a different $\Delta_h$, $f(\alpha)$ and $\beta^*(\alpha)$ for every area now. This is why the conditions for a change in choice within every area have to be considered accordingly. We obtain

| Option 1 chosen $(u_1 > u_2)$ | Option 2 chosen $(u_1 < u_2)$ | Value of $\beta^*(\alpha)$ |
|---|---|---|
| $\beta > \beta^*(\alpha)$ ; $\mathbb{A}_1$ | $\beta < \beta^*(\alpha)$ | $\beta^*(\alpha) < 0$ |
| $\beta < \beta^*(\alpha)$ ; $\mathbb{A}_2 \vee \mathbb{A}$ | $\beta > \beta^*(\alpha)$ | $1 < \beta^*(\alpha)$ |
| $\beta < \beta^*(\alpha)$ ; $\mathbb{B}$ | $\beta > \beta^*(\alpha)$ ; $\mathbb{C}$ | $0 < \beta^*(\alpha) < 1$ |

**Table 6** *Overview about the areas, choices and values of $\beta^*(\alpha)$ for conditions from (4.11)*

In order to visualize the last table, we get

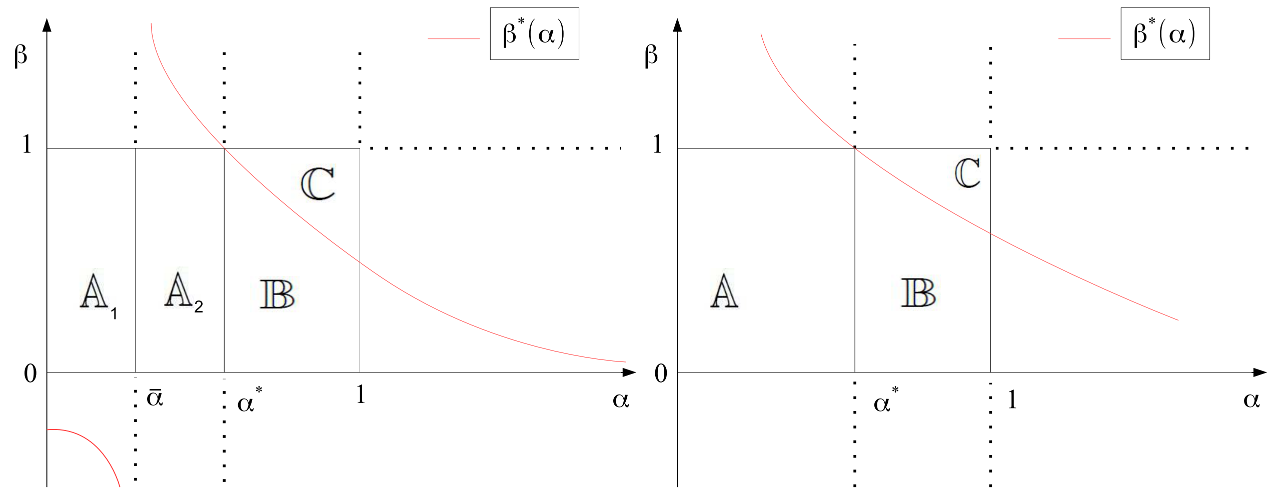



**Figure 5** *Sketch of $\beta^*$ for case (4.2) subcase b) with $\bar{\alpha} \in (0,1)$ (left-hand side) and with $\bar{\alpha} \notin (0,1)$ (right-hand side)*

Since we are still in trade-off case 1, we know that individuals are happier with Option 1 for an $\alpha < \alpha^*$. Thus, in the beginning, we are located in area $\mathbb{A}_1$, where we obtain a $\beta^*(\alpha)$ which is below 0. We know from the conditions above that $\beta$ needs to be larger than $\beta^*(\alpha)$

for Option 1 to be preferred in terms of choice, which is always the case as we only allow for strictly positive parameters. Put differently, if individuals are happier with Option 1, there is no $\beta$ we allow for that would make individuals choose Option 2. With an increasing $\alpha$ we enter area $\mathbb{A}_2$, where we face a $\beta^*(\alpha)$ that is above 1. This indicates that a $\beta$ larger than $\beta^*(\alpha)$ is a possibility we do not allow for. Hence, for a strictly positive $\beta$, Option 1 will always be preferred.

As we pass $\alpha^*$ individuals feel happier with Option 2. For areas $\mathbb{B}$ and $\mathbb{C}$, we can make clear distinctions for which $\beta$ Option 1 is preferred as $\beta^*(\alpha)$ is between 0 and 1. Given the conditions above, if $\beta$ is smaller than $\beta^*(\alpha)$, Option 1 will be chosen ($\mathbb{B}$), as the weight on happiness within the utility function is sufficiently low such that the direct effect of $x$ and $y$, which is larger for Option 1 here, is emphasized more strongly than the happiness effect from Option 2. Otherwise, Option 2 yields a higher utility ($\mathbb{C}$), as a $\beta$ above the cutoff puts enough weight on happiness within the utility function such that individuals are not only happier but also choose Option 2.

It might also be the case that $\bar{\alpha}$ is not in the unit interval. This is especially the case when the differences $\Delta_x$ and $\Delta_y$ are rather 'far' apart. We then only consider two sets of $\alpha$-values and three areas, i.e. $\mathbb{A}$ and $\mathbb{B}$ or $\mathbb{C}$. Eventually, we can draw the following proposition:

**Proposition 3** *Given definition 2 and that conditions (4.11) for trade-off (4.1) subcase b) hold, if weights $\alpha$ and $\beta$ take on values such that we are in area*

$$\begin{bmatrix} \mathbb{A} \vee \mathbb{A}_1 \wedge \mathbb{A}_2, \\ \mathbb{B}, \\ \mathbb{C}, \end{bmatrix} \textit{ utility from Option 1 is always } \begin{bmatrix} \textit{larger} \\ \textit{larger} \\ \textit{smaller} \end{bmatrix} \textit{ than utility from Option 2.}$$

**Proof.** See appendix. ■

Ultimately, the difference between cases (4.1) subcase a) and b) is determined by the numerator of $\beta^*(\alpha)$. It is either positive describing the former case or it is negative describing the latter.

Applying the same intuition as before to trade-off case 2, we can distinguish (4.2) subcase a) $(1-\gamma)\Delta_y - \gamma\Delta_x > 0$, and (4.2) subcase b), i.e. $(1-\gamma)\Delta_y - \gamma\Delta_x < 0$. We obtain similar results compared to (4.1) subcases a) and b). This originates in the signs of the numerator, which in the case of (4.1) and (4.2) for subcase a), respectively, subcase b) are the same.

# 5 How do people make decisions?

## 5.1 Do people choose what makes them happy?

Finally, do people choose what makes them happy? We consider again:

| $h \backslash u$ | $u_1 > u_2$ | | $u_1 < u_2$ | |
|---|---|---|---|---|
| $h_1 > h_2$ | coincide | I | contradict | II |
| $h_1 < h_2$ | contradict | III | coincide | IV |

**Table 7** *Utility and Happiness*

When do people behave in a consistent way? If we bring the propositions for happiness and utility together, we can set up the following theorems indicating when happiness and choice coincide. As we consider *two* different trade-offs with each trade-off featuring *two* subcases, we can draw *two* theorems.

We begin with:

**Theorem 1** *If $\alpha > \alpha^*$ and $0 < \beta(\alpha) < 1$ or if $\alpha < \alpha^*$ and $\beta > \beta^*(\alpha)$, then happiness and choice will coincide*
*(i) in trade-off 1 if and only if $(1-\gamma)\,\Delta_y + \gamma\Delta_x > 0$,*
*(ii) in trade-off 2 if and only if $(1-\gamma)\,\Delta_y + \gamma\Delta_x < 0$.*

**Proof.** See appendix. ■

We express the second theorem:

**Theorem 2** *If $\alpha > \alpha^*$ and $\beta > \beta^*(\alpha)$ or if $\alpha < \alpha^*$ and $0 < \beta(\alpha) < 1$, then happiness and choice will coincide*
*(i) in trade-off 1 if and only if $(1-\gamma)\,\Delta_y + \gamma\Delta_x < 0$,*
*(ii) in trade-off 2 if and only if $(1-\gamma)\,\Delta_y + \gamma\Delta_x > 0$.*

**Proof.** See appendix. ■

So ultimately, we can determine parameter line-ups for which happiness and choice coincide given our framework. But can we explain every decision an individual makes?

## 5.2 Can our model explain an individual's decision?

Starting with trade-off (4.1) subcase a), we can show that the combination $h_1 < h_2$ and $u_1 > u_2$ (i.e. combination III in Table 7) cannot be explained using strictly positive parameters as we cannot capture III with an appropriate area. The same holds for trade-off (4.2) subcase a). Thus, we observe that for subcases where the inequality between the payoff differences $\Delta_x$ and $\Delta_y$ in connection with $\gamma$ is in favor of the positive difference, which both subcases share, the numerator of $\beta^*(\alpha)$ emerges positive, causing combination III to be inexplicable. The difference between cases (4.1) subcase a) and (4.2) subcase a) arises from the different trade-offs[10]. Furthermore, we use the areas introduced and described above in section 4.2.2 to clarify intuitively what the model can explain. We get

| $h\backslash u$ | $u_1 > u_2$ | $u_1 < u_2$ |
|---|---|---|
| $h_1 > h_2$ | $\alpha < \alpha^*; \beta > \beta^*(\alpha)$; Area $\mathbb{A}$ | $\alpha < \alpha^*; \beta < \beta^*(\alpha)$; Area $\mathbb{B}$ |
| $h_1 < h_2$ | $\alpha > \alpha^*; \beta > \beta^*(\alpha) > 1 \vee \beta < \beta^*(\alpha) < 0$ | $\alpha > \alpha^*; 0 < \beta < 1$; Area $\mathbb{C}$ or $\mathbb{C}_1 \wedge \mathbb{C}_2$ |

**Table 8** *Parameter properties for case (4.1) subcase a)*

Regarding case (4.1) subcase b), we cannot explain combination $h_1 > h_2$ and $u_1 < u_2$ (i.e. combination II from Table 7). This also holds for (4.2) subcase b), except the relationship between $\alpha$ and $\alpha^*$ is different. For these two cases, we detect the inequality of payoff differences to be in favor of the negative difference, turning the numerator of $\beta^*(\alpha)$ negative. This leads $\beta^*(\alpha)$ to tend to a negative infinity for $\alpha = \bar{\alpha}$. If this occurs, combination II becomes inexplicable using standard economic theory. This is visualized in the following table. We get:

| $h\backslash u$ | $u_1 > u_2$ | $u_1 < u_2$ |
|---|---|---|
| $h_1 > h_2$ | $\alpha < \alpha^*; 0 < \beta < 1$; Area $\mathbb{A}$ or $\mathbb{A}_1 \wedge \mathbb{A}_2$ | $\alpha < \alpha^*; \beta > \beta^*(\alpha) > 1 \vee \beta < \beta^*(\alpha) < 0$ |
| $h_1 < h_2$ | $\alpha > \alpha^*; \beta < \beta^*(\alpha)$; Area $\mathbb{B}$ | $\alpha > \alpha^*; \beta > \beta^*(\alpha)$; Area $\mathbb{C}$ |

**Table 9** *Parameter properties for case (4.1) subcase b)*

**Proposition 4** *From the previous three Tables 7, 8 and 9, we know for strictly positive weights on happiness and utility (i.e. $0 < \alpha < 1$, $0 < \beta < 1$ and $0 < \gamma < 1$), that the model can always explain two combinations where happiness and utility coincide and one out of the two where they contradict each other, regardless which trade-off an individual faces.*

[10] The role $\alpha$ plays is for (4.2) subcase a) the exact opposite as for (4.1) subcase a).

Answering the question from the beginning of this section, our model extension can explain even more than standard utility theory suggests. If we set $\beta$ equal to zero, a parsimonious model emerges, such that utility is represented by a standard CES function. For this case, we can show when happiness and utility coincide. Using standard utility theory, we explain an average of 81 % of choices and preferences observed in terms of happiness for scenarios in Benjamin et al. (2012) representing trade-off case (4.1). This characterizes combinations which do not imply a difference between both concepts, i.e. they coincide.

Our model interprets happiness and utility as two equal concepts for $\beta$ being equal to one. In that case, our utility function could be characterized as 'hedonic utility.' Under the assumption of $\beta$ being in the unit interval, our model explains an average of 16 % points more of the beforehand mentioned empirical evidence, by also covering a combination where happiness and utility diverge. Ultimately, the question remaining is: What is a possible explanation for combination III for subcase a) and combination II of subcase b) which limits our model?

## 5.3 Model implications

Looking at the last two tables, we can describe three possible combinations. The trade-off case representing the sleep and income scenario is (4.1) subcase a). The combination we cannot explain is individuals being happier with Option 2, but actually choosing Option 1 (i.e. combination III). Apart from that, our model can rationalize 97 % of the empirical evidence.

Relaxing assumptions, we might be able to display every combination. There are two possible ways to achieve this.

### 5.3.1 Considering economic bads

It might generally be that certain scenarios feature payoffs which are usually considered economic goods, but which some people may perceive as economic bads. Therefore, the assumption that parameters must be strictly positive could be relaxed, for which we could describe all cases depending on which payoff is viewed as the economic bad, either one or both. Thus, going back to case (4.1) subcase a) without an $\bar{\alpha}$ between 0 and 1 (this can be done for every case possible), we would need to allow for values of $\beta$ larger than 1 to describe combination III (which can also be seen in Table 8). In terms of the model, nothing changes, except for when $\alpha$ is larger than $\alpha^*$. In particular, combination III and IV change due to allowing economic bads. Given the former, individuals who are happier with Option 2, i.e. more income and less sleep, decide for less income and more sleep, i.e. Option 1, and need a $\beta$ that is larger than $\beta^*(\alpha)$, which itself is larger than 1. This indicates that the direct effect of $x$ and $y$ is negative as $(1-\beta)$ becomes smaller than 0, indicating that the individual only cares about being happy and evaluates the rest, i.e. income and sleep, as essentially bad or not beneficial in terms of utility. However, as one might be able to make the case for sleep being an economic bad, it is unrealistic to describe income as something individuals prefer less of, i.e. view income as an economic bad.

### 5.3.2 Random decision-making

One could think of random decision-making as being a reason for the inexplicable combination II for cases (4.1) subcase b) and (4.2) subcase b) and III for cases (4.1) subcase a) and (4.2) subcase a). Evidence that individuals are conflicted between two different choices and, thus, decide randomly can be found in everyday life. Considering random decision aids, such as the coin flip or random decision-makers on the Internet, there are individuals who cannot make a decision for reasons that are manifold (e.g. analysis paralysis or lack of awareness).

This leads to the result that represents the least share of individuals (i.e. the 1 % from scenario 1) to be an outcome of random decision-making. Assuming a uniform distribution of individuals who are characterized by deciding randomly upon emotions (i.e. happiness) and rationales (i.e. utility), we find individuals within every combination making random decisions.

Knowing that the sleep vs. income scenario's inexplicable combination is III, the share of individuals maximizing their utility with Option 1, but not their happiness (i.e. they maximize happiness with Option 2) is 4 %. This stems from four possible combinations and 1 % of individuals representing this combination. If we want to compute and then compare this across all scenarios which fulfill the necessary conditions to reflect trade-off case (4.1) and subcase a), we have to make certain assumptions. Firstly, we need to identify the option-specific payoffs $x$ and $y$ for every scenario, quantify them carefully, such that we can classify every scenario, and compare it with the categorization of the trade-off cases from above. After identifying which scenario fits trade-off case (4.1), we can calculate the average of individuals whose decision can be traced back to be entirely random. Based on our calculations, an average of 2.8 % of people decide on combination III and 97 % on the other three[11]. Accordingly the share of individuals across all fitting scenarios deciding randomly is 11.2 %[12].

# 6 Conclusion

Subsequently, do people choose what will make them happy? There are cases where choice and happiness clearly differ. For these cases, we presented robust evidence from Benjamin et al. (2012), which we focused on explicitly to build our theoretical framework.

We showed a general model that explains choice-inferred utility and happiness for no trade-off and trade-off cases. Given the latter happiness and utility coincide if the parameters satisfy certain properties. The analysis of the values describing the option indifference regarding happiness, i.e. $\alpha^*$, and utility, i.e. $\beta^*(\alpha)$, is essential. As for the latter, an altering of $\alpha$ will change the value of utility indifference accordingly, which is why its properties were described extensively.

Ultimately, given the empirical evidence observed for scenarios fitting trade-off (4.1), standard utility theory explains an average of 81 %. Our model adds 16 % points, making us able to describe an average of 97 % of empirical evidence. We suggest the remaining percentage to be an outcome entirely originating from random decision-making. Assuming a uniform distribution, a share of 11.2 % of individuals is characterized by deciding randomly, whereas our model covers the share left, i.e. 88.8 % of individuals.

[11]This does not add up to an even 100 % as descriptive percentages from Benjamin et al. (2012) of one scenario do not add up to 100 %.

[12]Scenarios reflecting our trade-off case 1 are scenario 1, 4, 9 and 10 from the Cornell study and scenario 1, 4, 11 and 13 from the Denver study. Since scenarios 9, 10 and 11 do not include any payoffs, they are omitted from further calculations. The combination we cannot explain using our model is III. Thus, calculating the average and assuming a uniform distribution leads to presented 11.20 %. The calculation itself begins with adding up the percentages from combination III of the scenarios above, is then multiplied by four due to assuming a uniform distribution and, subsequently, divided by the number of scenarios we are considering, i.e. five.

## References


Adhvaryu, A., J. Fenske, and A. Nyshadham (2019): "Early Life Circumstance and Adult Mental Health," Journal of Political Economy, 127(4), 1516–1549.

Aghion, P., U. Akcigit, A. Deaton, and A. Roulet (2016): "Creative Destruction and Subjective Well-Being," American Economic Review, 106(12), 3869–97.

Benjamin, D. J., O. Heffetz, M. S. Kimball, and A. Rees-Jones (2010): "Do people seek to maximize happiness? Evidence from new surveys," Discussion paper, National Bureau of Economic Research.

Benjamin, D. J., O. Heffetz, M. S. Kimball, and A. Rees-Jones (2012): "What Do You Think Would Make You Happier? What Do You Think You Would Choose?.," American Economic Review, 102(5), 2083 – 2110.

Bentham, J. (1996): The collected works of Jeremy Bentham: An introduction to the principles of morals and legislation. Clarendon Press.

Campante, F., and D. Yanagizawa-Drott (2015): " Does Religion Affect Economic Growth and Happiness? Evidence from Ramadan ," The Quarterly Journal of Economics, 130(2), 615–658.

Deaton, A. (2018): "What do self-reports of wellbeing say about life-cycle theory and policy?," Journal of Public Economics, 162, 18 – 25, In Honor of Sir Tony Atkinson (1944-2017).

Easterlin, R. (1974): Does Economic Growth Improve the Human Lot? Some Empirical Evidencepp. 89–125. Nations and Households in Economic Growth: Essays in Honour of Moses Abramowitz, David, P.A. and Reder, M.W. (eds), Academic Press, New York and London.

Easterlin, R. A. (1995): "Will raising the incomes of all increase the happiness of all?," Journal of Economic Behavior & Organization, 27(1), 35–47.

Fliessbach, K., B. Weber, P. Trautner, T. Dohmen, U. Sunde, C. E. Elger, and A. Falk (2007): "Social comparison affects reward-related brain activity in the human ventral striatum," science, 318(5854), 1305–1308.

Frey, B. S., and A. Stutzer (2000): "Maximising happiness?," German economic review, 1(2), 145–167.

Frey, B. S., A. Stutzer, et al. (2002): "The economics of happiness," World economics, 3(1), 1–17.

Fudenberg, D., and D. K. Levine (2006): "A Dual-Self Model of Impulse Control," American Economic Review, 96(5), 1449 – 1476.

Gul, F., et al. (1991): "A theory of disappointment aversion," Econometrica, 59(3), 667–686.

Kahneman, D., and A. Deaton (2010): "High income improves evaluation of life but not emotional well-being," PNAS, 107(38), 16489–16493.

Kahneman, D., P. P. Wakker, and R. Sarin (1997): "Back to Bentham? Explorations of Experienced Utility," The Quarterly Journal of Economics, 112(2), 375–405.

Kimball, M., and R. Willis (2006): "Utility and happiness," University of Michigan, pp. 1–67.

Liberini, F., M. Redoano, and E. Proto (2017): "Happy voters," Journal of Public Economics, 146, 41 – 57.

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

——— (1986): "Disappointment and dynamic consistency in choice under uncertainty," The Review of Economic Studies, 53(2), 271–282.

Luttmer, E. F. (2005): "Neighbors as negatives: relative earnings and well-being," The Quarterly journal of economics, 120(3), 963–1002.

O'Donoghue, T., and M. Rabin (1999): "Doing It Now or Later," American Economic Review, 89(1), 103–124.

Oswald, A. J., E. Proto, and D. Sgroi (2015): "Happiness and Productivity," Journal of Labor Economics, 33(4), 789–822.

Rayo, L., and G. S. Becker (2007): "Evolutionary efficiency and happiness," Journal of Political Economy, 115(2), 302–337.

Stevenson, B., and J. Wolfers (2008): "Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox," (14282).

Weiss, A., T. C. Bates, and M. Luciano (2008): "Happiness is a personal (ity) thing: The genetics of personality and well-being in a representative sample," Psychological Science, 19(3), 205–210.