Is It Possible to Raise National Happiness?

Alberto Prati[1] and Claudia Senik[2]

[1]UCL, LSE and University of Oxford; United Kingdom.
[2]Sorbonne University; Paris School of Economics; IUF; France

December 18, 2024

Abstract

We revisit the famous Easterlin paradox by considering that the Cantril scale of life evaluations refers to a changing context, hence is itself regularly reinterpreted. We propose a simple model of rescaling based on both retrospective and current life evaluations. We apply it to reconstruct an index of latent happiness and find that the well-being of Americans has substantially increased, on par with GDP, health, education, and liberal democracy, from the 1950s to the early 2000s. We also shed light on other happiness puzzles, including the apparent stability of life evaluations during COVID-19, or the absence of parental happiness.

Keywords: Happiness, Life Satisfaction, Subjective Well-Being, Easterlin Paradox, Cantril Ladder, Rescaling, Gallup, SOEP

JEL Codes: I31

Acknowledgments

We are grateful for financial support from the French National Research Agency, through the Investissements d'avenir framework ANR-17-EURE-0001; from the AAPG2021(WRKCOV19); from CEPREMAP; from the John Fell Fund; and from the Oxford Wellbeing Research Centre. We benefited from the remarks and suggestions of participants in seminars in Bordeaux University, Frei University (Berlin), Luxembourg University, the London School of Economics, the Paris School of Economics, Keio University, University College London, and at the Conference of European Statistics Stakeholders, as well as discussion with Farid Anvari, Roland Bénabou, Andrew Clark, Jan-Emmanuel De Neve, Mark Fabian, Marc Fleurbaey, Caspar Kaiser, Carol Graham, Andrew Oswald, Nicolas Jacquemet, Leonardo Pejsachowicz, Leandro Prados de la Escosura, Marco Ranaldi. We would like to thank the Roper Centre for its support with the historical US records and Rachel Penrod for her support with the more recent U.S. records, as well as Gallup.Inc and Wellbeing for Planet Earth for making the global data collection possible. Rémy Bellaunay and Isaac Parkes provided excellent research assistance. All errors are our own.

Is it possible to raise national happiness?

"Raising the incomes of all does not increase the happiness of all. This is because the material norms on which judgements of well-being are based increase in the same proportion as the actual income in the society" (Easterlin, 1995, p.44).

[. . .] *people typically think that they were worse off in the past and will be better off in the future, although their reports on present happiness remains constant over time"* (Easterlin, 2001, p.472).

Policymakers and statistical agencies are increasingly recognizing life satisfaction as a valuable tool to monitor the well-being of the population. Among the OECD countries that have a national wellbeing framework, 85% use life satisfaction (Mahoney, 2023). In the UK, public policy appraisals are encouraged to assess costs and benefits in terms of their impact on life satisfaction (HM Treasury, 2021). If life satisfaction is used as metric of national happiness, a natural question arises: is it possible to raise national happiness?

The answer is not trivial, because of a well-known puzzle in the happiness literature: the so-called Easterlin paradox, i.e. the empirical observation that, over the long run, say a couple of decades of economic growth, the trend in self-declared life satisfaction (LS) is flat. This is called a paradox for two reasons: first, it is at odds with what most people would expect; second, it is not consistent with the lessons of cross-section and cross-country analysis, where the richer are also happier.[1] The implications of the Easterlin paradox are considerable. If "raising the income of all" does not (durably) "increase the happiness of all" (Easterlin, 1995) then we should change our way of thinking about economic decision-making and, more generally, the purpose and organization of our societies and governments.

Another long-standing stylized fact, recognized by Richard Easterlin (Easterlin, 2001) and corroborated by several studies (e.g., Diener and Oishi, 2000; Hagerty, 2003; Busseri and Samani, 2019; Prati and Senik, 2022; Kaiser, 2022), is that people underestimate their past happiness and tend to be over-optimistic about their future happiness. "In every country, in every age group, from 18-29 to 50 and over, respondents, on average, rated their future happiness higher and their past happiness lower, with only a few trivial exceptions. [...] but in fact, over the entire period present happiness was, on average, constant. Thus, we have another paradox to explain – why people typically think that they were worse off in the past and will be better off in the future, although their reports on present happiness remains constant over time" (Easterlin, 2001, pp. 471-472).

Hence, Easterlin's stylized facts to be explained are: (i) the flatness of the average LS series in

[1] The Easterlin paradox applies also when we broaden the scope of human development beyond economic growth. When the Human Development Index (HDI) is considered instead, the paradox holds: over the long term, the HDI grew significantly in many countries without them resulting any happier; yet, at a point in time, happier countries tend to have a higher HDI (Yin et al., 2023). We replicate the cross-sectional evidence of the positive relationship between life satisfaction and economic affluence (measured as log GDP per capita) and human development (measured as HDI) in the appendix, figure A1.

the long run, despite economic growth; and (ii) the disconnect between current and remembered satisfaction with a systematic underestimation of past life satisfaction.

But can one really consider the chronicle of life satisfaction as reflecting the dynamics of people's *latent* life satisfaction? The question arises because of the way this concept is measured. Large surveys ask people to report their level of happiness, or life satisfaction, on a bounded scale, usually a 0-10 scale. However, obviously, the units of the scale do not correspond to any objective observable magnitude. Unlike meters or kilograms, the steps of the scale do not refer to any unit (Kaiser and Oswald, 2022b); rather they are referred, e.g. by the initial Cantril formulation, as "the worst possible life" (0) and "the best possible life for you" (10).[2] Therefore, when asked to score their life satisfaction on a scale, respondents do not report known information such as their weight or height, but rather a relative level of happiness, i.e., their own position in the total scope of possibilities, spanning the entire realm of misery and bliss in the concrete world they are living in. In other words, self-reported levels of life satisfaction are context-dependent, they are time and space dependent.

If one must decide whether "raising the incomes of all [will] increase the happiness of all?", this remark is crucial. Easterlin's question indeed asks about the dynamic evolution of *latent, experienced* happiness. The question to be decided is then the following: What is relative and context-dependent: latent experience or the measurement scale? What is the underlying mechanism that keeps the happiness trend flat: *hedonic treadmill* or *rescaling*?

Tenants of the *hedonic treadmill* interpretation, following Richard Easterlin, posit that latent satisfaction itself is relative, i.e. subject to adaptation and relative concerns. This rejoins Kahneman and Tversky's prospect theory (1979), whereby latent experienced happiness depends not on the absolute level (say, of income) enjoyed by a person, but on the gap with respect to a benchmark level. That benchmark, in turn, can be the level that one has been enjoying in the past and has modified their level of aspirations. It can also be formed by observation of the situation of other people (say, reference income). The fact that intertemporal judgments do not match concurrent judgments can be explained as a memory illusion: When asked to reconstruct their past satisfaction, respondents confound their past circumstances with their past feelings (Easterlin, 2001).

By contrast, partisans of *rescaling* have pointed out that the happiness scale is naturally context-dependent because it is bounded. "The 'best possible life for you' is a shifting standard that will move upwards with rising living standards", as put by Deaton (2008, pp.12-13). An improved world of possibilities will then make reporting styles more stringent, and a person will require a higher level of latent satisfaction in order to choose a given step on the life satisfaction scale. Daniel Kahneman (2009) has also discussed this possibility and, although starting from the opposite prior, came to the conclusion that "citizens of different countries do not adapt to their level of prosperity", therefore leaning more towards the hypothesis of rescaling rather than the hedonic treadmill. *Rescaling* offers another explanation for why intertemporal judgments do not match the chronicles of life satisfaction: When reconstructing their past satisfaction, people may recall their past satisfaction correctly, but they measure it on their current scale, which differs from the one used at the epoch.

[2]Here is a picture of a ladder. Suppose we say that the top of the ladder represents the best possible life for you and the bottom represents the worst possible life for you. Where on the ladder do you feel you personally stand at the present time?

Formally, the fundamental difference between the two conjectures is that the hedonic treadmill assumes a time-dependent utility function, while the rescaling assumes a time-dependent reporting function. In the former, the context acts on a person's actual *latent* satisfaction, whereas in the latter, it only acts on their interpretation of the scale. In the former, people become harder to satisfy as the world improves; in the latter, they become more satisfied, but tougher in their grading style.

Both interpretations are observationally equivalent, and to date, no study has been able to definitely depart them, in spite of many attempts. We believe there is little reason to consider the two conjectures - hedonic treadmill or rescaling - as mutually exclusive and, in this paper, we do not try to quantify the role of each of them. Rather, we focus on the second one and show what a reconstructed time series of life satisfaction would be like under the rescaling hypothesis, based on a new method that exploits current and retrospective satisfaction. Most of the existing literature has interpreted life satisfaction time series from the perspective of hedonic treadmill. Here, we explore it from the perspective of rescaling. We provide four main contributions.

First, we develop a method to correct for rescaling. We identify rescaling based on the discrepancy between retrospective satisfaction and the satisfaction reported at the epoch, while also accounting for memory errors. We provide some clear instructions to survey data stakeholders about how to use this method using existing or future longitudinal data.

Second, we 'reconstruct' the chronicle of latent life satisfaction in the United States in the long term. For this, we put together the longest time series of comparable life satisfaction scores ever published, starting with the original interviews run by Cantril in 1959. We show that, depending on the assumption about the type of rescaling process, we obtain a life satisfaction index that evolves more or less closely in parallel with GDP per capita and the Human Development Index. The Easterlin paradox, therefore, disappears.

Third, when we apply our life satisfaction index in times of crisis, some other happiness puzzles disappear. For example, life satisfaction time series around the world were surprisingly flat during the COVID-19 pandemic in 2020 (Helliwell et al., 2022), when more than half of the world's population was under lockdown.[3] Instead, our new method shows that 2020 was a year of substantial rescaling, during which aspirations contracted and happiness declined. Similarly, in Ukraine, the average level of life satisfaction remained unchanged in 2023 compared to 2018. However, our index reveals that the Russian invasion had a significant negative impact on the life satisfaction of Ukrainians - an effect that remains hidden when rescaling due to declining aspirations is overlooked.

Finally, we run a battery of psychometric tests to assess the validity and reliability of our new measure. We collect new data both from an online panel, using Prolific.com, and from a socio-economic panel, in collaboration with German SOEP. Different panel datasets concur that, when rescaling is taken into account, observable life events increase the explained variance in individual happiness (the R squared are 1.5 to 2 times higher). The signs of the regression coefficients are consistent when using either LS or our new index as dependent variables, except the one associated with childbirth, which flips to a positive value when accounting for rescaling.

[3]Life satisfaction from national panel data display a flat trend too. This can be observed in the UK Understanding Society Survey panel, in the Household Income and Labour Dynamics panel in Australia, and in the German Socio-Economic Panel.

This paper belongs to the literature discussing key assumptions for the analysis of self-reported satisfaction data (Bond and Lang, 2019; Kaiser, 2022; Fabian, 2022a; Kaiser and Oswald, 2022b; Benjamin et al., 2023b,a; Andreoni et al., 2024). Arguably, the papers closest to ours are Kaiser (2022) and Benjamin et al. (2023b). Kaiser (2022) is the first to propose using retrospective satisfaction reports to correct for rescaling of happiness time series. Crucially, his correction is based on *ordinal* reports of subjective changes ("Would you say that you are more satisfied with life, less satisfied, or feel about the same as you did a year ago?"), while we rely on *cardinal* reports of past satisfaction levels, which allow us to estimate the size of the rescaling. Benjamin et al. (2023b) uses a similar conceptual framework to model scale differences, and proposes some calibration questions to adjust for it, in a similar vein to our paper. However, it studies differences in scale use *across respondents*, while we study it *over time*.

The remainder of the paper is organized as follows. Section 1 briefly discusses the state of the literature on happiness scales. Section 2 discusses the different assumptions underlying hedonic treadmill and rescaling, and provides a formal framework to identify rescaling using retrospective satisfaction data. Section 3 introduces the several datasets that we analyze, and section 4 presents the results. Section 5 describes a series of exercises to assess the psychometric properties of our new life satisfaction index, and provides some instructions to survey data stakeholders. Section 6 concludes.

1 Literature

The question about how people use happiness scales has naturally been asked and studied since the infancy of this literature. Bernard van Praag and the authors of the Leyden school ran the first experiments asking whether people were able to evaluate their welfare position on a scale. Their ultimate objective was to estimate the relationship between income and welfare (well-being). They asked the so-called Income Evaluation Question, that takes the following form (with variants): *"While keeping prices constant, what after-tax total monthly income would you consider for your family as: very bad... bad... insufficient... sufficient... good... very good"*. They showed that the best way for a respondent to provide information on their welfare function is to choose the answers in such a way that each of the six levels corresponds to a jump of $1/6$ in utility. This is the so-called *equal quantile assumption*, admittedly "a comeback to the cardinal utility concept" (Van Praag, 1971, p.388). More generally, they showed that when asked to evaluate any quantity on a scale, even if the verbal descriptions associated with the steps of the scale are vague, respondents will tend to interpret the question as if the steps of the proposed scale corresponded to an equal partition of the substance referred to (Van Praag, 1991; Van Praag and Frijters, 1999). For example, Van Praag et al. (1988) ran an experiment asking subjects to connect age levels to subjective labels that refer to other adults such as *young, somewhat young, middle-aged, somewhat old, and old*. Again, subjects spontaneously divided the maximum life expectancy of the moment into equal quantities, i.e. five intervals of 15 years.

In the following decades, a series of experiments reached the conclusion that the life satisfaction

scale can be interpreted as cardinal, with equidistance or linearity of scale steps i.e. that when people answer the life satisfaction question, the difference they make between 3 and 4 is the same as the difference between 7 and 8 (Krueger and Schkade, 2008; Oswald, 2008; Kaiser and Oswald, 2022b,a; Kaiser, 2022). In a recent paper, Kaiser and Vendrik (2020) discussed this issue at length and concluded that the evidence indicates that participants tend to perceive scales containing more than three response items as approximately linear.

In sum, the lesson from these experiments is that when asked to use a scale, people equate the maximum quantity with the highest step and the minimum with the lowest end of the scale and associate the intermediate steps with equal quantiles of the given substance. People "use the response scale to indicate the quantile of one's life in a reference distribution of possible lives" (Kaiser, 2022, p.412). Following these observations, here we assume that when asked to report their subjective well-being on a bounded scale, people realize the same projection exercise and interpret the scale as divided into equal quantiles of life satisfaction, equating the maximum level of latent happiness with the highest step (10) and the minimum with the lowest step (0), thus creating a "social standard of happiness" (Van Praag and Frijters, 1999). Note that the Cantril ladder actually explicitly suggests this interpretation.

Before us, rescaling, i.e. shifts and stretches of the life satisfaction scale, have been discussed and tested.[4] In a recent paper, Kaiser (2022) reported a number of previous experiments showing evidence that people calibrate response scales more or less "stringently" depending on the benchmark they use, in particular, when using their memories of the past. In the same spirit, Howard and his coauthors compared the self-reported evaluations of an episode in real time and retrospectively, and interpreted any difference as denoting a scale change (Howard et al., 1979a,b; Howard and Dailey, 1979). Similarly, Köke and Perino (2017) underlined the inconsistency between regressions of life satisfaction on observed versus reported changes in several demographic characteristics, and attribute these differences to the change in respondents' reporting function. They show that the scale becomes more stringent when respondents earn higher incomes. However, these authors recognize the difficulty of distinguishing a shift in the scale from a shift in actual satisfaction.

This last remark also applies to the "preference drift" uncovered by the Leyden school. This expression refers to the change in people's aspirations as their own income increased (Van Praag, 1971), where income aspirations are measured by people's answer to the Minimum Income Question.[5] Stutzer (2004) used Swiss data and showed that the coefficient on this aspiration level is negative in the regression of life satisfaction. More recently, Stillman et al. (2015) compared successful and unsuccessful applicants to a migration lottery in New Zealand, and found that despite significant increases in real incomes, successful migrants did not report higher well-being scores than unsuccessful applicants, but did report a higher position compared to their pre-immigration status,

[4]Some authors (e.g. Fabian, 2022b) use the term "scale norming" to refer to what we call here "rescaling". Other authors (e.g. Kahneman and Krueger, 2006) talk about "aspiration treadmill" to refer to a specific type of rescaling, where the top of the scale shifts upward over time. The concept of rescaling discussed here is similar, but not equivalent to the one used in the literature on social welfare functions (see, e.g. Adler, 2019, pp.42-43).

[5]"What household income per month would you consider an absolute minimum in order to make ends meet and without running into debt even if you reduce your needs to a minimum? We do not only mean housekeeping allowance but all essentials, including insurance, rent, taxes and so on."

a contradiction that the authors attribute to rescaling.

However, again, these results can be interpreted either as reflecting hedonic treadmill or rescaling. The flat trend in self-reported life satisfaction can reflect an actual stagnation of latent well-being as the level of aspirations increases, or alternatively, a more stringent reporting style (Frederick and Loewenstein, 1999).

To date, the concept of rescaling over time has been used mainly in the health literature, where the most common approach, known as "then-test" (Howard et al., 1979a; Howard and Dailey, 1979; Howard et al., 1979b; Schwartz and Sprangers, 2000), relies on a method similar to the one we develop here. This literature has focused on how quality of life scores change over time in people with serious health conditions. Rescaling could help explain why partial adaptation is observed after becoming disabled (Oswald and Powdthavee, 2008).[6] For instance, people who were victims of spinal cord injury have been showed to use a quality-of-life scale quite differently in the five years following the accident (Schwartz et al., 2018).[7] The seminal study of this strand of literature is Brickman et al. (1978), which showed that people who became paraplegic after an accident rate their life *before* the injury as significantly higher than a control group of people who did not endure an accident. In the same line, Van Leeuwen et al. (2012) studied current and past life evaluations over time in persons with spinal cord injury, and found substantial rescaling using a "comparison score" between current life satisfaction and life satisfaction before the accident.

Because we are interested in the dynamics of subjective well-being, we have discussed the literature related to the potential rescaling process over time, i.e. the diachronic deformation of the way people interpret the happiness scale. An ancient and abundant literature in psychology and social sciences has dealt with the synchronic heterogeneity in the way people interpret happiness scales or report their subjective assessments in general, sometimes referred to as "differential item functioning". The "anchoring vignette" literature is an example of such efforts (King et al., 2004; Kapteyn et al., 2007). Clearly, the questions and methods imagined to address the two issues are close. A recent paper by Benjamin et al. (2023b) for example, proposes and applies a new method to quantify and correct the degree of heterogeneity across demographic groups of the population in the way they use subjective well-being scales. Our approaches are complementary.

In the next section, we sketch out a simple framework that clarifies the different assumptions underlying the models of hedonic treadmill and rescaling. Based on this framework, we propose a method for studying the dynamics of latent satisfaction based on either model.

[6]The widely cited idea that victims of severe accidents quickly and completely recover their happiness scores has actually little empirical support and is rejected by Oswald and Powdthavee (2008). This is also true for the (often misquoted) cross-sectional study by Brickman et al. (1978). Brickman and colleagues conclude that "the results for accident victims appear to be less supportive of adaptation level theory. The accident victims did not tend to take more pleasure in ordinary events and rated themselves significantly less happy in general than controls" (p.921). The confusion might come from the sentence which immediately follows "It should be noted, however, that the paraplegic rating of present happiness is still above the midpoint of the scale and that the accident victims did not appear nearly as unhappy as might have been expected" (p.921).

[7]Schwartz et al. (2018) follows the approach developed by Oort et al. (2005) and measure a change in the use of the scale by studying the change in conditional correlations between life satisfaction and observables variables over time. This method has the advantage of avoiding memory errors, but conflates the effects of rescaling (using the scale more/less stringently) and reprioritization (valuing some things more and some things less).

2 Models of latent and reported life satisfaction

We first start with a simple presentation of the two competing models of the dynamics of latent and nominal life satisfaction over time: the hedonic treadmill and rescaling models. We show how both of Easterlin's paradoxes can be accounted for by each model. Building on this, we propose a method to reconstruct the time series of latent life satisfaction within the rescaling framework.

2.1 Easterlin's hedonic treadmill versus rescaling

We start with Easterlin's first stylized fact. Easterlin (1995) proposes a model in which latent satisfaction h derives from the difference between people's actual level of income, say (c), and their level of aspiration, say ($\bar{c}$). Then his conjecture is that the time trend of life satisfaction remains flat due to an adaptation ("habit formation") process whereby income aspirations $\bar{c}$ rise together with people's achievements c. Because satisfaction comes from the difference between achievements (c) and aspirations ($\bar{c}$), when both magnitudes increase roughly at the same pace, latent satisfaction h remains unchanged.[8]

Easterlin also mentions another possible reason for the flatness of the LS curve over time, namely that satisfaction depends on comparisons with other people's lot (another interpretation of $\bar{c}$), so that on average, if the lot of everyone improves, the average happiness remains unchanged (zero-sum game at best, negative-sum game if comparisons are only upward).[9]

Concerning the second stylized fact, that people think they were less happy in the past and they will be happier in the future, Easterlin's interpretation is again based on the hedonic treadmill hypothesis (Easterlin, 2001, p.465): "People think they were less happy in the past and will be happier in the future, because they project current aspirations to be the same throughout the life cycle, while income grows. But since aspirations actually grow along with income, experienced happiness is systematically different from projected happiness". An important point is that this interpretation implies that, when trying to answer the question about their past satisfaction, people do not try to recall their past level of satisfaction h_{t-1}, but rather think about their past income c_{t-1} and judge what their satisfaction must have been on that ground, given their current higher aspirations.[10]

Note that Easterlin does not make a distinction between latent life satisfaction felt by people (h) and their self-declared nominal level of satisfaction on the proposed scale (LS). He plots the evolution of life satisfaction over time and takes the (flat) observed pattern at face value. Implicitly, this

[8] "Material aspirations increase commensurately with income, and as a result, one gets no nearer to or farther away from the attainment of one's material goals, and well-being is unchanged" (Easterlin, 2003). "People [...] project current aspirations to be the same throughout the life cycle, while income grows. But since aspirations actually grow along with income, experienced happiness is systematically different from projected happiness. Consequently, choices turn out to be based on false expectations" (Easterlin, 2001, p.44).

[9] "Put generally, happiness, or subjective well-being, varies directly with one's own income and inversely with the incomes of others. Raising the incomes of all does not increase the happiness of all, because the positive effect of high income on subjective well-being is offset by the negative effect of higher living level norms brought about by the growth in incomes generally" (Easterlin, 1995, p.36).

[10] Easterlin explicitly writes: "When asked about well-being five years ago or five years hence, a person [...] can be thought of as telling us how she would feel today if she had the income y_m (worse off) or y_3 (better off)." (Easterlin, 2001, p.474).

suggests that there is something inherent to the human psyche or biology that makes it impossible to increase latent happiness beyond a given level, say 10 on a 0-10 scale. Set-point theories of happiness (Lykken and Tellegen, 1996) are consistent with this view, although they are not explicitly endorsed by Richard Easterlin.

An alternative assumption to the "hedonic treadmill", is the idea that shifting aspirations deform the reporting scale, as described by Kahneman and Krueger (2006, p.16): "if people gradually adjust their aspirations to the utility that they normally experience, an improvement of life circumstances would eventually lead them to report no higher life satisfaction than they did before, even if they were experiencing higher utility than previously. In this scenario, experienced utility could rise even while one's global evaluation of life satisfaction remained constant".

The difference in the two views is the following. Both are interested in the dynamic evolution of latent happiness h_t over a time span $t \in [0, T]$. One can only observe the chronicle of nominal life satisfaction LS_t which is self-declared in each period t. Suppose that, in a given country, on average, LS_0 and LS_T are equal, e.g. a "representative agent" declared a level of 7 in the two dates, although income, educational achievement and health have gone through a trend of positive growth.

According to the hedonic treadmill interpretation, the flat trend in LS is due to an adaptation process, whereby:

$$LS_t = f(h(c_t, \bar{c}_t)) \tag{1}$$

where $f(.)$ is the reporting function of the feeling h into the nominal life satisfaction scale, and $h(.)$ depends on $\bar{c}_t$, i.e., the level of aspirations of the agent, which increases at a similar pace as c.

By contrast, the rescaling view considers that aspirations $\bar{c}_t$ impact the act of *reporting*, but not the latent satisfaction itself, i.e.:

$$LS_t = f(h(c_t), \bar{c}_t) \tag{2}$$

Equations (1) and (2) might appear similar, but they describe two very distinct processes. The crucial difference is that, according to the hedonic treadmill view, life circumstances and the reference point $\bar{c}$ jointly determine satisfaction: these two aspects are non-separable. Therefore, a change in the reference point implies an intrinsic change in feelings. By contrast, according to the rescaling view, life circumstances are mapped into latent satisfaction first, and this latent satisfaction is subsequently reported according to the reference point $\bar{c}$. Therefore, the well-being effect of a change in personal circumstances is separable from a change in the reference point. Said differently, hedonic treadmill assumes a time-dependent utility function, while rescaling assumes a time-dependent reporting function. This is, in a nutshell, the difference between hedonic treadmill and rescaling.

In the rescaling framework, an increase (or a drop) in latent satisfaction can be masked by a change in aspirations. When answering the life satisfaction question, people have to project their latent satisfaction on the proposed 0-10 scale, which is described as the scope of possibilities spanning from "the worst possible life" (h^{min}) and "the best possible life for you" (h^{max}). As this scope changes over time, so does the signification of a given step of the scale. Hence, latent satisfaction h can be rising although nominal satisfaction LS remains on a flat trend.

2.2 A reverse engineering exercise

In order to address the rescaling issue, based on a meta-analysis, Hagerty (2003) recommends collecting inter-temporal judgements of life satisfaction as collected in the same point of time (e.g. current, past and long past life satisfaction, as declared in a given date): "[...] analyzing *intertemporal* changes in happiness may require different measurement methods from analyzing *concurrent* differences across people in the same time period. Economists make a similar distinction by adjusting "current" dollars to "inflation-adjusted" dollars before making intertemporal comparisons. For similar reasons, we may need to adjust concurrent measures of happiness for inflation of standards over time" Hagerty (2003, p.136, italics in the original).

An important point is that, to explore this conjecture, one has to assume that people remember, at least partly, their past satisfaction, and can judge the extent to which they are happier or less happy than they used to be. We will discuss this point at length later on, in section 2.7.

The inflation metaphor is all the more useful, as life satisfaction is measured on a scale without any unit, which means that its dynamic evolution over time is ideally represented as an index, with reference to a baseline period (as the index of real GDP growth, for instance). In order to calculate the desired index of latent satisfaction h over time, one must do a reverse engineering exercise and correct the levels of nominal life satisfaction LS for the rescaling process due to the projection. In this section we explain how.

The rescaling view involves two potential mechanisms that characterize a change in the reporting function: scale shifting and scale stretching (or shrinking), with k_t the shift parameter and d_t the stretch parameter such that:

$$k_t = h_t^{min}; \;\; d_t = \frac{h_t^{max} - h_t^{min}}{LS_{max} - LS_{min}} = \frac{h_t^{max} - h_t^{min}}{10} \tag{3}$$

Where h_{min} and h_{max} are, respectively, the worst and best possible life. Hence, at any time t, an individual will declare a level of life satisfaction:

$$LS_t = h_t \times \frac{1}{d_t} - \frac{k_t}{d_t} \tag{4}$$

This equation is key. It describes the mapping of the (unbounded) feeling h_t into the (bounded) 0-10 satisfaction scale. The parameters d_t and k_t regulate *scale stretching* and *scale shifting*, respectively. As the index t shows, k and d can change over time, as the boundaries of latent happiness (h_{min} and h_{max}) change. Our objective is to define the trajectory of h relative to a fixed point in time (a numeraire). Without loss of generality, we define this reference point as the reported life satisfaction (LS) at the first period of study ($t = 0$), i.e.:

$$h_0 \equiv LS_0 \tag{5}$$

So that:

$$k_0 = 0; \;\; d_0 = 1 \tag{6}$$

Then, latent satisfaction at any period t is:

$$h_t = LS_t \times d_t + \sum_{n=0}^{t}(k_n - k_{n-1}) \tag{7}$$

2.3 Using Retrospective Satisfaction

Obviously, with two unknown parameters k_t and d_t and unobservable h_t, it is impossible to decide between the rescaling and the hedonic treadmill interpretations. Both will generate the same time series of LS_t.

However, we can mobilize an additional source of information when it is available. This source is the level of past life satisfaction as reported in time t: $RLS_{t,t-1}$ (life satisfaction in $t-1$ as remembered in period t), for example: "How satisfied were you a year ago with your life?" (Socio-Economic Panel, 2023) or "Please think about a picture of a ladder. Suppose that the top of the ladder represents the best possible life for you, and the bottom represents the worst possible life for you. If the top step is '10' and the bottom step is '0'. On which step would you say you stood five years ago?" (Cantril, 1965). Retrospective satisfaction brings about a very useful property, namely it is expressed on a common scale with LS_t. That is, we can define $RLS_{t,t-1}$ as past latent satisfaction h_{t-1} reported on today's scale:

$$RLS_{t,t-1} = \frac{h_{t-1}}{d_t} - \frac{k_t}{d_t} \tag{8}$$

LS_t and $RLS_{t,t-1}$ are readily comparable and contain some relevant information on the variation of latent life satisfaction h between periods $(t-1)$ and t. We interpret the difference between LS_{t-1} and $RLS_{t,t-1}$ as due to rescaling and use this information to convert h_t and h_{t-1} into comparable units:

$$h_t = h_{t-1} + (LS_t - RLS_{t,t-1}) \times d_t \tag{9}$$

Based on this definition, we obtain the following recursive form to express latent satisfaction h at any time t:

$$h_t = LS_0 + \sum_{n=1}^{t}(LS_n - RLS_{n,n-1}) \times d_n \tag{10}$$

We reconstruct the index of subjective life satisfaction by connecting the points $h_0, ..., h_t$ defined as above.

We now have three unknown magnitudes (h_t and two unobserved parameters of potential rescaling: k_t and d_t), and two identifying equations (LS_t and $RLS_{t,t-1}$). This is not sufficient to readily identify h_t. To get around this identification problem, a number of restrictions are possible. In what follows, we will consider two mutually exclusive scenarios: one where the scale can shift, and one where the scale can expand or shrink. We then reconstruct the time series of h_t under each identifica-

tion restriction. As we will see, using simultaneously $RLS_{t,t-1}$ and LS_{t-1} in order to calculate d_t or k_t generates qualitatively similar time series, regardless of which identification restriction is adopted. Moreover, the no-shift and no-stretch scenarios represent, respectively, a lower- and upper-bound of latent satisfaction under some plausible conditions. Details of the model and of the conditions are in the appendix, section A4.[11]

2.4 No shift scenario

In this scenario, we assume that the bottom of the scale (worst possible life) is fixed over time and we set: $k_t = k_0 = 0$, $\forall t$, meaning that we consider the worst possible life as of period 0 as the benchmark. The scope of possibilities can then grow ($d > 1$) or shrink ($d < 1$), and the series of reconstructed h_t will write: $h_t = d_t \times LS_t$. Therefore:

$$h_t = LS_t \times \prod_{n=1}^{t} \frac{LS_{n-1}}{RLS_{n,n-1}} \tag{11}$$

2.5 No stretch scenario

Alternatively, we assume that the entire scale can shift, upward or downward, implying that the bottom and top of the scale shift in the same direction at the same pace, so that $d_t = d = 1$, $\forall t$. In this scenario, the difference between the best possible life and the worst possible life remains constant, but the set of possible levels of h can shift upward if life becomes better, for instance if life expectancy increases, or it can shift downward, in case of a war for instance.

The reconstructed time-series will then write:

$$h_t = LS_t + \sum_{n=1}^{t} (LS_{n-1} - RLS_{n,n-1}) \tag{12}$$

Authors such as Fabian (2022b) focus on this scenario, and some recent medical literature studying quality-of-life among paraplegics indeed suggest substantial scale shifting but no scale stretching/shrinking (Schwartz et al., 2018).

2.6 Raising the happiness of all or raising aggregate happiness

Notice that in both cases of no shifts or no stretch, the equations that describe the dynamics of latent happiness h_t are linear, so they apply to individual happiness as well as to the average happiness of a country's inhabitants. If we had long *panel* data containing information about current and remembered happiness, we could compute an index of reconstructed latent happiness for each individual. This would allow us to take into account individual circumstances, such as the

[11] The conditions are: (i) when people feel better than yesterday, their idea of best possible life must improve at least as fast as their idea of worst possible life; (ii) when people feel worse than yesterday, their idea of worst possible life must deteriorate at least as fast as their idea of best possible life.

relative position of each person in the general distribution of income. However, for the moment, the long time-series that we use are not panel, as will be described in the data section, so that we are restricted to studying the evolution of aggregate latent satisfaction, here average satisfaction. Accordingly, our index refers to a "representative" or aggregate agent at the country level. We will also analyze *short*-run individual panel data in section 5.1.

2.7 Memory correction and the M-LINE

So far, we have assumed away memory errors. At the aggregate level, this could be correct if memory errors had a structure of white noise, i.e. were randomly centered around zero. This practical assumption, however, is an unrealistic one. Previous studies suggest that most people are subject to a structured recall bias, in particular, they recall their satisfaction level as *lower* than it was (Prati and Senik, 2022). This raises the suspicion that the difference between retrospective satisfaction and life satisfaction reported in the past ($RLS_{t,t-1} \neq LS_{t-1}$) could be due to memory errors rather than rescaling.

To inquire, we run experiments under the identifying assumption that the memory bias tends to be constant. Indeed, although recall accuracy decreases over time (as predicted by the law of recency), the literature on memory retention does not support the hypothesis that a digit is more likely to be misremembered either upwards or downwards, and the most common models of recall accuracy define only the variance - and not the mean - of recall error as a function of time (Anderson et al., 1998; Kahana, 2012). Accordingly, we assume memory errors to follow a stochastic process with a mean $\bar{m}$, which is potentially different from zero, and a variance v, which depends on the retention interval, i.e. $m_{t,t-n} \sim D(\bar{m}, v_n)$, where D is a symmetrical probability distribution.

In order to disentangle the role of memory biases and rescaling, the key of the identification strategy is to look at retrospective satisfaction over a very short time interval. On the one hand, rescaling tends to be a slow cumulative process, where distortions of the reporting functions pile up over time but are negligible over small intervals (except in the case of important collective shocks, e.g. a war). On the other hand, as soon as respondents are asked to evaluate their past life, the bias associated with the recall function kicks in.

To estimate the recall bias, we set up an online longitudinal survey using Prolific.com, where we interview a representative sample of the American population twice, two weeks apart. After considering attrition and exclusion restrictions, 1,647 respondents were eligible for the analyses. Section A3.1, in the appendix, provides more information on this data collection.

We interview the same participants twice, at two week distance, and ask them (1) in Wave 1 (January 3, 2024), to evaluate their life using the 0-10 Cantril ladder (LS_{t-2w}); (2) in Wave 2 (January 17-20, 2024), to evaluate their life as of two weeks ago, using the retrospective 0-10 Cantril ladder ($RLS_{t,t-2w}$). We find that retrospective satisfaction is, on average, *lower* than current satisfaction, even for that very small time interval ($LS_{t-2w} > RLS_{t,t-2w}$). The size of the underestimation represents the memory bias parameter, $\bar{m}$, which is estimated to be 0.12 ($p - value < 0.001$; 95% CI: [0.6;0.17]). This estimate is very similar to the one obtained by Anvari et al. (mimeo), who

ran a similar exercise with respondents from various countries. In their data, the two-week apart average difference between LS and RLS in Western countries is 0.13 ($LS_{t-2w} - RLS_{t,t-2w}$=0.13; $p - value < 0.001$; 95% CI: [0.8;0.19], N=1,332).[12]

Based on our external calibration exercise, we thus adjust our index of reconstructed latent happiness and define a new index that we call **M-LINE** (**M**emory-adjusted **LI**fe satisfaction **N**oetic **E**valuation).[13]

In case of shift and no stretch (with $d_t = 1, \forall t$ and $k_0 = 0$), the reconstructed latent happiness at time t will be:

$$h_t = LS_t + \sum_{n=0}^{t-1} LS_n - \sum_{n=1}^{t} RLS_n - t\bar{m}. \tag{13}$$

In case of no shift (with $d_0 = 1$ and $k_t = 0\ \forall t$), the reconstructed latent happiness at time t will be:

$$h_t = LS_t \times \prod_{n=1}^{t} \frac{LS_{n-1}}{RLS_{n,n-1} + \bar{m}} \tag{14}$$

We will refer to latent satisfaction calculated under the assumption of no shift (eq. (13)) as M-LINE$^{k=0}$. For latent satisfaction with no stretch (eq. (14)) we will use M-LINE$^{d=1}$. Finally, latent satisfaction under the assumption of no stretch *and* no shift is simply nominal life satisfaction, LS.

3 Data

3.1 Long-run time series

We base our long-run calculation of the M-LINE on the Cantril ladder. Compared to other measures of life satisfaction, the Cantril ladder tends to elicit thoughts related to wealth (Nilsson et al., 2024), thereby making the Easterlin paradox even more salient. Besides, this is the only life satisfaction measure that is collected around the globe every year, and constitutes the main variable that is analyzed and discussed in the annual World Happiness Report.

Our exercise requires historical records on both life satisfaction and retrospective life satisfaction, measured on a cardinal scale. Both measures were originally introduced by Cantril (1965). The first poll was run by Cantril's team in 1959, and continued by Gallup inc. up to today. However, to date, no study has been able to reconstruct the evolution of the Cantril ladder from its invention to present day. To do so, we used recently digitalized archival data from the Roper Centre at Cornell University, and merged them with unexplored data from Gallup US polls in early 2000s, and with the Gallup World Poll.

[12] These results should not be extrapolated to non-Western countries. In South-Africa, i.e. the only large non-Western country sample that they collect, Anvari et al. (mimeo) find that the two-week apart average difference between LS and RLS is much larger, i.e. 0.44 ($LS_{t-2w} - RLS_{t,t-2w}$=0.46; $p - value < 0.001$; 95% CI: [.29;.62], N=293).

[13] The adjective "noetic" refers to a *judgment* (life satisfaction) made about a *representation* (past life), associated with *semantic memory* (not the memory of a particular episode, but of life in general).

The time series starts with the original micro-data collected by Hadley Cantril in 1959 and was published six years later in *The pattern of human concerns*. Between 1959 and 2000, Gallup Inc. ran ten nationally representative surveys in the US that asked about both the Cantril ladder and the retrospective ladder. Between 2001 and 2005, Gallup Inc. collected micro-data on both the Cantril ladder and the retrospective ladder annually, via its "Mood of the nation" module of the U.S. Social Series survey. Between 2006 and 2008, Gallup Inc. asked these same questions in the "Gallup World Poll". The time series ends in 2009 when the retrospective ladder was removed from the World Poll, only to be reintroduced in 2023. Overall, we obtain a pseudo-panel of representative cross-sections of the U.S. population from 1959 to 2008.

Table 1: Datasets for the long-term analyses

Source	Year	LS		RLS		LS-RLS		N
		mean	st.err.	mean	st.err.	mean	st.err	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Cantril	1959	6.58	0.06	5.89	0.07	0.69	0.07	1,502
U.S. Polls	1964	6.81	0.05	5.92	0.06	0.89	0.06	1,550
U.S. Polls	1971	6.56	0.05	5.76	0.06	0.80	0.07	1,549
U.S. Polls	1972	6.42	0.06	5.48	0.07	0.93	0.08	1,245
U.S. Polls	1974	6.45	0.07	5.45	0.08	1.00	0.08	1,201
U.S. Polls	1976	6.67	0.07	5.66	0.08	1.01	0.08	1,052
U.S. Polls	1979	6.48	0.05	5.72	0.06	0.76	0.07	1,568
U.S. Polls	1985	6.42	0.05	5.75	0.06	0.67	0.06	1,545
U.S. Polls	1987	6.54	0.03	5.79	0.04	0.75	0.04	4,206
U.S. Polls	1989	6.44	0.04	5.96	0.05	0.48	0.05	2,033
U.S. Polls*	1998	7.10	-	5.70	-	0.40	-	-
U.S. Social Series	2001	6.96	0.06	5.95	0.07	1.01	0.08	993
U.S. Social Series	2002	6.70	0.06	6.16	0.07	0.54	0.07	1,007
U.S. Social Series	2003	6.75	0.06	6.13	0.07	0.62	0.08	988
U.S. Social Series	2004	6.60	0.06	6.11	0.07	0.49	0.08	1,001
U.S. Social Series	2005	6.67	0.06	6.26	0.07	0.41	0.07	998
World Poll	2006	7.18	0.06	6.15	0.08	1.03	0.08	995
World Poll	2007	7.51	0.05	6.58	0.06	0.93	0.06	1,212
World Poll	2008	7.28	0.06	6.77	0.06	0.51	0.07	994

Survey-weighted estimated means based on representative samples of the US population.
*Only toplines available in 1998.

Table 1 reports the summary statistics of life satisfaction and retrospective life satisfaction. The first measure of interest is the evolution of nominal life satisfaction, reported in column (3). The Easterlin paradox predicts that we should observe no increase in nominal life satisfaction over the long term; there is, however, a 0.7-point difference in nominal life satisfaction as measured in 1959 and 2008. This difference should be interpreted with caution, since the two moments of progression in the time series (i.e., the end of the 1990's and 2005) correspond to a moment of change in sources (from U.S. Polls to U.S Social Series and World Poll). If we take these discontinuities into account,

in the spirit of Stevenson and Wolfers (2008), no clear trend appears *within* each of the three sub-periods.

Column (7) in table 1 displays the average difference between current and retrospective life satisfaction reported at the same point of time, in the same survey wave. As appears clearly by checking different years, the difference is systematically positive and, actually, quite large (0.74 scale points, on average). This perceived improvement is, however, not accompanied by a corresponding increase in current life satisfaction as reported over successive years. Hence, the data do exhibit both of Easterlin's paradoxes.

To construct the M-LINE, we aggregate the data into 5-year periods (i.e., from 1959-1963 to 2004-2008) and assume a uniform distribution within each period (so that the average LS of a period is the average of the observed years). We end up with a 10-period dataset, where, for each period t, we observe current satisfaction LS_t and retrospective satisfaction for the previous period, $RLS_{t,t-1}$. The resulting dataset has enough information to identify either the shift or stretch parameters, and subsequently allow us to calculate the M-LINE under alternative assumptions.

In constructing the M-LINE, we calibrate it on a constant memory bias $\bar{m} = 0.12$, and consider the two (no-shift and no-stretch) scenarios based on the identification restrictions detailed in sections 2.4 and 2.5. As a robustness check, we also repeat the analyses with a memory bias parameter ranging from 0 (no memory bias) to 0.6 (5 times the estimated bias), and report the results in the appendix, figures A2-A8.

3.2 Short-run times-series (cross-section)

We use two longitudinal cross-sections, where a representative sample of the population was interviewed multiple times.

The first one is CAMME, a representative quarterly survey of the French population, run by the French National Statistical Institute. From 2016 onward, participants have been asked not only about their current level of satisfaction with life (0-10), but also where they feel they stood about one year ago (0-10). We will analyze this dataset in section 4.2.

The second one is the Gallup World Poll, an annual survey representative of 99% of the world's adult population, where the Cantril ladder (0-10) is asked. We leverage a new question on retrospective Cantril ladder that was added to the Gallup World Poll questionnaire in 2023, and that was fielded in Ukraine. Also, this data set will be analyzed in section 4.2.

3.3 Short-run times-series (panel)

Panel data provide the ideal condition for measuring rescaling. We use four national panels for this purpose.

The first panel is the German Socio-Economic Panel (SOEP), a well-studied high-quality panel dataset. In the first few years of the panel, between 1984 and 1987, participants from West Germany were asked to report not only their level of life satisfaction (0-10) but also their retrospective life

satisfaction one year ago (0-10), thereby allowing us to calculate the M-LINE. This part of the data set has already been used by Odermatt and Stutzer (2019, Appendix S7) and Klemm (2022). Before our study, this was the only extant large panel data that measured RLS. We will analyze this data in section 5.3.

The second panel was also provided by SOEP. Each year, SOEP opens a submission call for researchers to propose new questions to be included in the SOEP Innovation Sample (SOEP-IS). We were able to include the following question in the panel: "All things considered, how satisfied with your life were you in each of the last five years?". The question was fielded in 2022, so that participants reported their past life satisfaction (0-10) as of 2017, 2018, 2019, 2020 and 2021. The panel structure allows to compare the retrospective answers given in 2022 ($RLS_{2022,t}$) with the chronicle of nominal life satisfaction reported in each year t (i.e., LS_t). We will analyse it in section 4.2.

The third panel is an online sample, representative of the American population, that was interviewed twice, at two-week distance, in January 2024. We briefly describe this panel in section 2.7. It is a balanced panel of 1,628 respondents who were asked about their life satisfaction using the Cantril ladder (0-10) and their retrospective life satisfaction two weeks ago and five years ago. They were also asked about their perception of rescaling. The interviews were conducted on Prolific.com. More details are available in the appendix, section A3.1. Hereafter, we will refer to this dataset simply as the "US online panel".

The fourth and last panel is an online panel of U.K. residents who were interviewed at one year distance. They were asked to report current LS in 2023 and 2024, and reconstructed LS ("how satisfied were you in 2023?") in 2024, on a continuous 0-to-100 scale. They were also asked to provide information about how certain they felt with their answers - a question that we use to assess the subjective reliability of self-declared satisfaction - and a direct question about comparison: "Compared to one year ago, how satisfied with your life are you nowadays?", measured on a continuous scale from -100 to +100. Interviews were conducted on Prolific.com. About two-thirds of the 1,502 participants who were surveyed in February 2023 responded again one year later, in February 2024.[14] This leaves us with a balanced panel of 1,005 observations. More details are available in the appendix, section A3.2. Hereafter, we will refer to it simply as "U.K. online panel", to distinguish it from the "U.S. online panel".

Table 2 gives an overview of the various datasets that we use in our analyzes of the M-LINE on the short run.

4 Results

4.1 Half a century of well-being growth in the USA (1959-2008)

Figure 1 illustrates the evolution of the M-LINE, where each of the two panes corresponds to one scenario. In both figures, the vertical arrows indicate the position and the extent of the scale,

[14] 13 participants who did not pass an attention check were dropped from the analysis.

Table 2: Datasets for the short-term analyses

Data Source	Country	Years	Data Structure	Repres.	Nb.interviews
CAMME	France	2016-2023	cross-section	✓	53,918
Gallup World Poll	Ukraine	2018, 2023	cross-section	✓	1,954
SOEP core	West Germany	1984-1987	panel	✓	27,345
SOEP-IS	Germany	2017-2022	panel	✓	6,742
U.S. online panel	USA	2024	panel	✓	3,256
U.K. online panel	United Kingdom	2023-2024	panel		1,974

estimated on the basis of LS_t, LS_{t-1} and $RLS_{t,t-1}$. The blue line indicates the evolution of nominal satisfaction LS_t. The green dashed line indicates the level of satisfaction of the same corresponding year that will be reported in retrospect at the next period, i.e., $RLS_{t+1,t}$.

The fact that the blue line is above the green line (to an extent that cannot be explained by memory bias) suggests that the scale has moved upward. Finally, the red line, indicates the reconstructed evolution of latent satisfaction when accounting for rescaling, i.e., the M-LINE.

Each figure also shows the evolution of two objective well-being statistics, i.e. income (as measured by real GDP per capita in 2015 USD), and human development (as measured by the Augmented Human Development Index, or AHDI). The AHDI (Prados de la Escosura, 2021) offers a long-term historical view of human development, in the spirit of the HDI, which, instead is available from 1990 only. It includes the following measures, combined as an equally weighted geometric average: life expectancy at birth; years of schooling; liberal democracy index; real GDP per capita.

Each figure has two vertical axes. The y-axis on the left of the graph is expressed in latent life satisfaction points of the 1959 scale. The y-axis on the right is expressed as an index with base 1959.

Between 1959 and 2008, GDP per capita in the US almost tripled, and the AHDI increased by 50%. When comparing these statistics with the stagnation of nominal life satisfaction, the Easterlin paradox appears in all its evidence. Should one conclude that this progress was futile?

Taking rescaling into account leads to a different conclusion. Although the precise estimate of the M-LINE depends on the underlying assumptions, both images exhibit an important progression of latent well-being. Aspirations (represented by the top-end of the scale) have substantially increased too, which explains how Americans could at the same time feel better over time but report a steady level of nominal satisfaction.

In figure 1, top pane, we assume that the bottom of the scale is anchored at zero ($k = 0$). The trajectory of latent satisfaction, represented by the M-LINE, follows an upward and slightly convex trend, like GDP per capita. The end-point, in 2008, is more than 2.5 times as high as the starting-point, in 1959. Said differently, if we could measure satisfaction in the USA in 2008 using an unbounded satisfaction scale with the same step size as the 1959 scale, the average answer would be 17.2. This is a markedly different dynamic from what nominal satisfaction would suggest. The interpretation of the top of the scale changed dramatically, too. We estimate that the latent satisfaction associated with reporting $LS = 10$ was almost two and half times as high in 2008 as it

was in 1959 ($d_{1959} = 1; d_{2008} = 2.42$). As a consequence, 7/10 (the modal response to the Cantril ladder question) had a very different meaning in 2008 than 50 years earlier.

The bottom pane of figure 1 displays the evolution of latent well-being in the case the scale can shift but not stretch/shrink over time ($d = 1$). In this case too, the M-LINE displays a substantial increase of latent satisfaction, that roughly doubled between 1959 and 2008. The assumption that the scale simply shifted upward (without any stretch) may not be very plausible,[15] but it is a useful exercise to decompose the effects of the stretching and shifting parameters. In this scenario, where the size of the scale is fixed, the increase in latent satisfaction is entirely driven by the scale shift by 6 points ($k_{1959} = 0; k_{2008} = 5.57$). In the appendix, tables A1-A2 report the data underpinning figure 1, while figures A2-A8 repeat the analyses with a memory bias parameter spanning from 0 (no memory bias) to 0.6 (5 times the estimated bias) and show that, even in the most conservative scenario, latent life satisfaction is clearly increasing.

4.2 The M-LINE during dire times: the case of the pandemic in Europe and of the war in Ukraine

The previous section showed that, when rescaling is taken into account, the chronicle of latent life satisfaction in the USA is following a similar trend as GDP per capita or the HDI, instead of the flat trend in nominal life satisfaction. This is due to the correction of the difference between contemporary and (the underscored) retrospective evaluations. But is the past systematically undervalued or can the correction go the other way? In other words, is the M-LINE always above the curve of nominal satisfaction or can it be below?

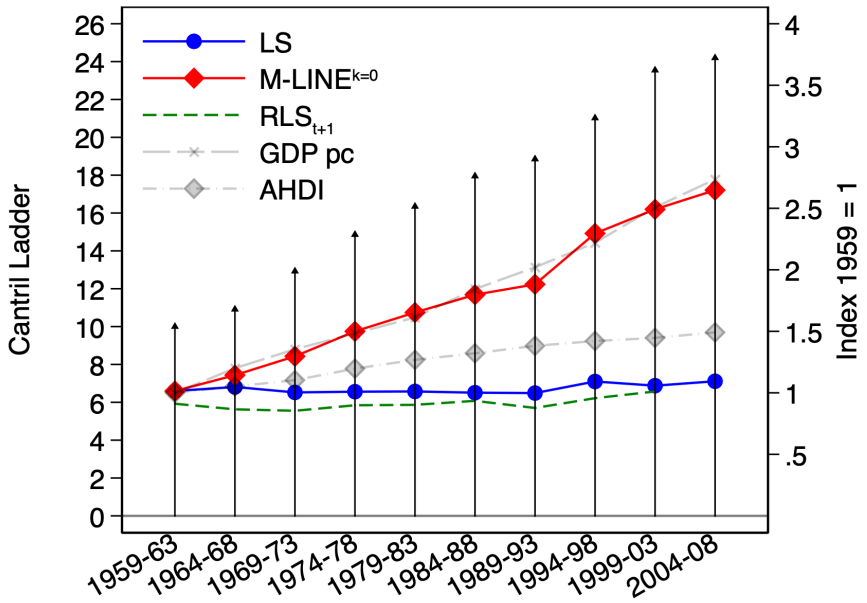
To find out, we study the evolution of the latent scale during dire times, using three datasets. The first allows reconstructing *a posteriori* the dynamics of latent satisfaction during the COVID-19 pandemic in Germany. The second one leverages retrospective life satisfaction questions asked during the pandemic in France. Third and last, we estimate the change in scale use among residents of Ukraine after the Russian invasion, and recalculate the effect of the war on their latent satisfaction. Tables A3-A5, in the appendix, summarize the results presented graphically in this section.

4.2.1 The pandemic in Germany

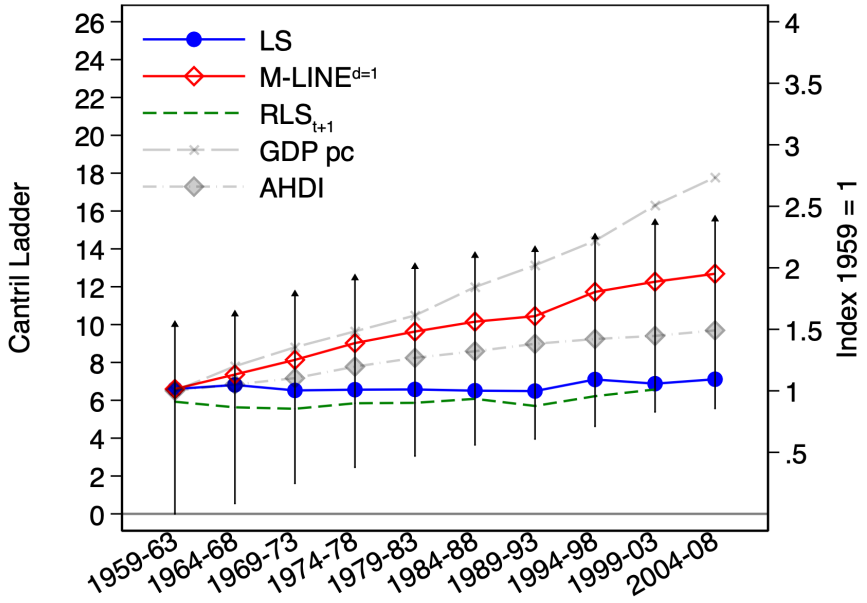
In Germany, nominal life satisfaction during the pandemic increased. Data from the World Happiness Report suggest a 0.2-point increase in 2020 compared to the prepandemic period (Helliwell et al., 2021), and Germany was not an isolated case, as nominal life satisfaction in 2020 was statistically equal to or higher than in 2017–2019 in 75% of the countries surveyed. These figures are at odds with the statistics on mental health, which show that the incidence of major anxiety and depression disorders increased by about 25% worldwide during the pandemic (Daly and Robinson, 2022).

Can rescaling explain this paradox? To make progress, we used the German Socio-Economic Panel, a well-studied high-quality panel dataset. In 2022, the SOEP Innovation Sample (SOEP-

[15] This would imply that the latent satisfaction felt by many Americans in 1959 corresponds to $LS = 0$ on the 2008 scale.



(a) No shift



(b) No stretch

Figure 1: Latent satisfaction for different rescaling patterns (USA, 1959-2008)

Reading note: Panel (a) shows the M-LINE estimated under the no-shift hypothesis ($k = 0$). Panel (b) shows the M-LINE estimated under the no-stretch hypothesis ($d = 1$). LS, RLS and the M-LINE are measured on the y-axis of the left-hand side, expressed in life satisfaction points; GDP per capita and Augmented HDI are measured on the y-axis of the right-hand side, expressed as an index equal 1 in 1959. The upward arrows represent the life satisfaction scale adjusted for rescaling. LS question: "Think about a picture of a ladder. Suppose that the top of the ladder represents the best possible life for you, and the bottom represents the worst possible life for you. If the top step is "10" and the bottom step is "0", on which step of the ladder do you feel you personally stand at the present time?" [0-10]. RLS question: "On which step would you say you stood five years ago?" [0-10]. Data is aggregated in 5-year intervals. LS and RLS source: various datasets (cf table 1). GDP per capita source: St.Louis Federal reserve Bank. AHDI source: Prados de la Escosura (2021).

IS) asked its over 6,000 participants the following question: "All things considered, how satisfied with your life were you in each of the last five years?". The panel structure allows us to compare the retrospective answers given in 2022 ($RLS_{2022,t}$) with the chronicle of nominal life satisfaction reported between 2017 and 2020.[16]

On average, $RLS_{2022,t} + \bar{m} < LS_{t-1}$. As is most often the case, the score of remembered life satisfaction is below what people had declared in the corresponding year, even when we account for memory bias. However, the first year of the pandemic, i.e. 2020, stands out as an exception. Although the nominal level of life satisfaction reported in that year was *higher* than in the previous periods, the retrospective evaluations of that period indicate that it was significantly *worse* than previous years. Absent rescaling, this would mean that the first year of the pandemic was an exceptionally good year for the average German resident, whereas two years later, respondents were deluded in the recollection of a bad 2020. Conversely, in the framework of rescaling, this could mean that in 2020, the scale underwent some important contraction and/or downward shift, which inflated nominal satisfaction in spite of a lower latent satisfaction, as it appears in retrospective reports. Of course, we tend to be more convinced by the second interpretation.

We therefore reconstruct the time series of latent satisfaction for Germany over the years 2017-2020 and display the result in figure 2. As can be seen, because in that case all the retrospective evaluations are made in 2022, the dynamics of latent life satisfaction is reconstructed backward with 2022 as the baseline year of the index. In the figure, the M-LINE illustrates the trough in latent life satisfaction felt in 2020, that was not visible on the curve of nominal life satisfaction.

4.2.2 The pandemic in France

Although the reconstruction of the M-LINE in the German dataset was based on retrospective reports only, we would like to know if interviews *during* the pandemic would give similar results. Therefore, we move our focus to France and use the CAMME dataset, a representative survey of the French population, run by the French National Statistical Institute, which collects life satisfaction records within their consumer confidence survey. Each quarter, about 1,800 participants are asked not only about their current level of satisfaction with life, but also where they feel they stood about one year ago. CAMME collected retrospective life satisfaction before, during and after 2020, thereby providing an example how scale adjustment can be used "in real time".

In figure 3, the blue line shows the evolution of nominal life satisfaction measured in December each year between 2016 and 2021. If the latent scale was not affected by the pandemic, a 7/10 during lockdown, for example, would be equivalent to a 7/10 the year before and a 7/10 the year after. As the blue lines on figure 3 show, if there was no rescaling, then (i) there was no significant change in the life satisfaction of French residents between 2016 and 2023, and (ii) the "lockdown year", i.e. 2020, oddly stood out as one of the happiest years on record in France. When rescaling is considered instead, these conclusions are flipped. First, latent life satisfaction increased between 2016 and 2023, which was not immediately visible in the nominal data because the life satisfaction

[16] In 2021, SOEP-IS data collection was suspended, so we cannot make a comparison for that specific year. It started again in 2022.

scale underwent an expanding trend. Depending on the scenario, it either expanded by an annual average rate of 1.9%, or it shifted upward at an average annual rate of 0.12 scale points. Second, when rescaling is taken into account, 2020 turns out to be a year where French residents endured a fall of their life satisfaction, rather than a rise. In contrast to the positive trend of the previous years, in 2020 the scale is estimated to *contract* by 3% (instead of expanding), or to shift downward by 0.23 scale points (instead of upward). Arguably, this was due to a reduction in what the best possible life could have looked like in that year. The red lines in figure 3 show the M-LINE for each of our two scenarios. Regardless of the assumption, the M-LINE shows a recession of life satisfaction in 2020.

4.2.3 The war in Ukraine

The pandemic was a global phenomenon, but its effects on people's well-being were heterogeneous, and some citizens might have enjoyed the opportunities of those unusual times. Although no socio-economic event can be univocally considered to have a negative effect on wellbeing, war is likely the closest to fitting this definition. We therefore study the evolution of life satisfaction and rescaling during the ongoing war on Ukrainian territory.

We use the Gallup World Poll, an annual survey that is representative of 99% of the adult population of the world, including a representative sample of Ukrainian residents. We leverage a new question on retrospective satisfaction that was added to the Gallup World Poll questionnaire in 2023, where people were asked to measure their life five years ago on the Cantril ladder.

As illustrated by figure 4, self-reported levels of life satisfaction show no variation between 2018 and 2023 (blue lines). The reconstructed M-LINE (in red) offers a different conclusion, by revealing an important drop in latent happiness during the Russian invasion (2023) compared to pre-pandemic levels (2018). In terms of scale stretch (left pane), latent happiness fell by almost two steps on the initial 0-10 scale, which itself has contracted by more than two steps. In terms of scale shift (right pane), it appears that the scale has shifted downward by two steps, so that the "worst possible life" (0) as of 2023 would correspond to -2 on the initial 2018 scale (or conversely, the 0 "worst possible life" as of 2018 would be rated at 2 on the 2023 scale), and the "best possible life" (10) in 2023 corresponds to the 8th step on the 2018 scale. Incidentally, 2023 was a year of economic expansion in Ukraine, with a real GDP growth of 5.3% (IMF, 2024): the dissociation between the expansion of GDP and the contraction of the M-LINE in this context provides a clear indication that rescaling captures more than economic growth.

These three exercises show that the M-LINE does not automatically correct nominal scores of life satisfaction upward. In some cases, such as a social crisis associated with a pandemic or a war, rescaling suggests that the dip in latent satisfaction was deeper than what nominal satisfaction would suggest. In addition, these results suggest a unified solution to two recent puzzles of happiness economics, i.e., why life satisfaction data was flat during the pandemic and why Ukrainians report to be as happy during the war as before it. Previous literature has proposed that this may be due to the timing of the data collections (Easterlin and O'Connor, 2023, for the pandemic in Europe)

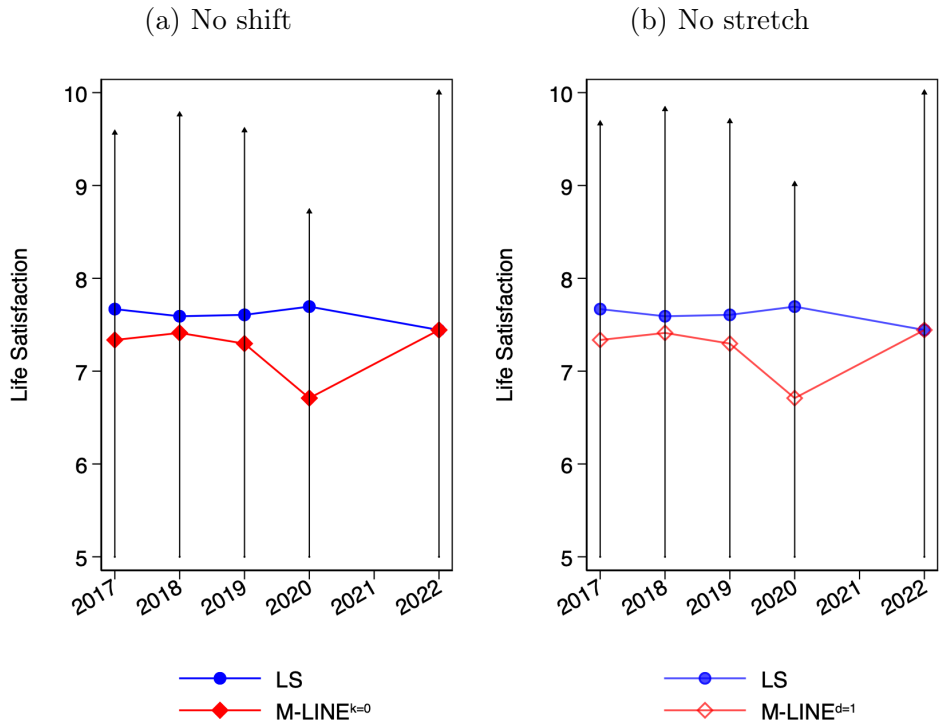


Figure 2: Wellbeing during the pandemic: M-LINE and nominal LS in Germany.

Reading note: Panel (a) shows the M-LINE estimated under the no-shift hypothesis ($k = 0$). Panel (b) shows the M-LINE estimated under the no-stretch hypothesis ($d = 1$). The upward arrows represent the life satisfaction scale adjusted for rescaling (truncated at 5/10 for better visual rendition). LS question: "How satisfied are you with your life, all things considered?" [0-10]. RLS question: "All things considered, how satisfied with your life were you in each of the following years? 2017, 2018, 2019, 2020" [0-10]. Source SOEP-IS. Data is from different individuals interviewed in a nationally representative survey of the German population in 2022.

or geographically heterogeneous effects (Coupe and Obrizan, 2016, for the war in Ukraine). We contribute to this debate by showing that both puzzles can be attributed - at least in part - to rescaling issues.

5 Validation

The soundness of our results depends on the quality of our new measure, the M-LINE. In what follows, we report several exercises aimed at assessing its psychometric validity and reliability, in comparison with current life satisfaction reports. For this, we use several data sources.

5.1 Construct validity with *objective* variables: does the M-LINE correlate with what is going on in people's lives?

The cornerstone of our validation exercise is to assess if the M-LINE correlates with important life events or change in people's objective life circumstances. For this, we need individual panel data. We run the analysis on two panels: (i) the first four waves of the German Socio-Economic Panel

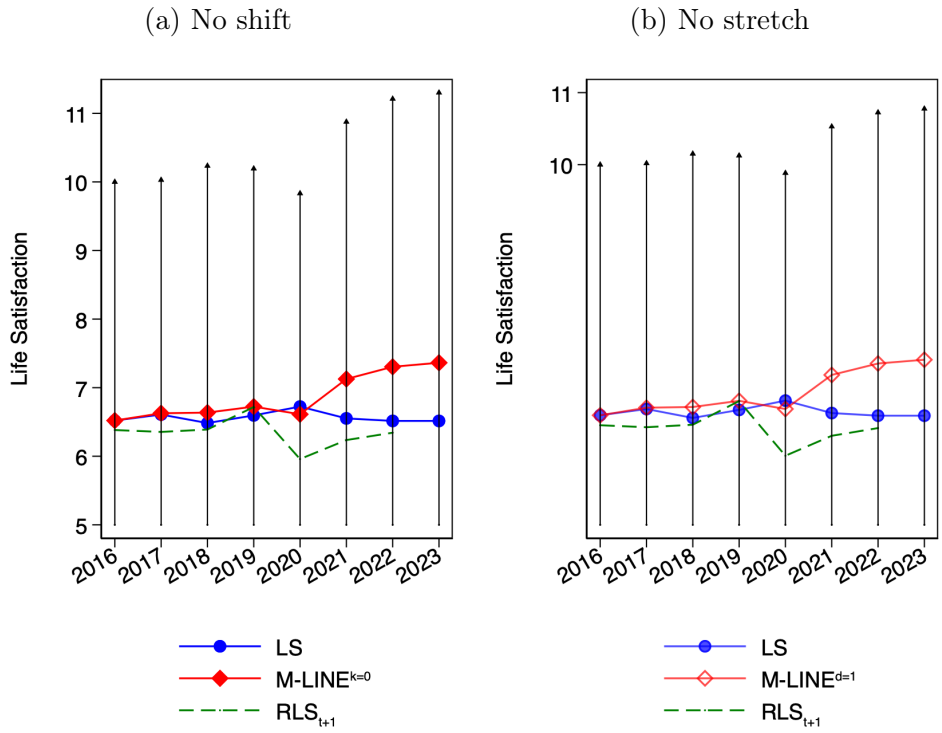


Figure 3: Wellbeing during the pandemic: M-LINE and nominal LS in France.

Reading note: Panel (a) shows the M-LINE estimated under the no-shift hypothesis ($k = 0$). Panel (b) shows the M-LINE estimated under the no-stretch hypothesis ($d = 1$). The upward arrows represent the life satisfaction scale adjusted for rescaling (truncated at 5/10 for better visual rendition). LS question: "Overall, how satisfied are you with the life you currently lead?" [0-10]. RLS question: "And when you think about last year, how did you place on a scale of 0 to 10?" [0-10]. Source CAMME. Data is from different individuals interviewed in 31 nationally representative surveys of the French population between 2016 and 2023, and aggregated in one-year intervals.

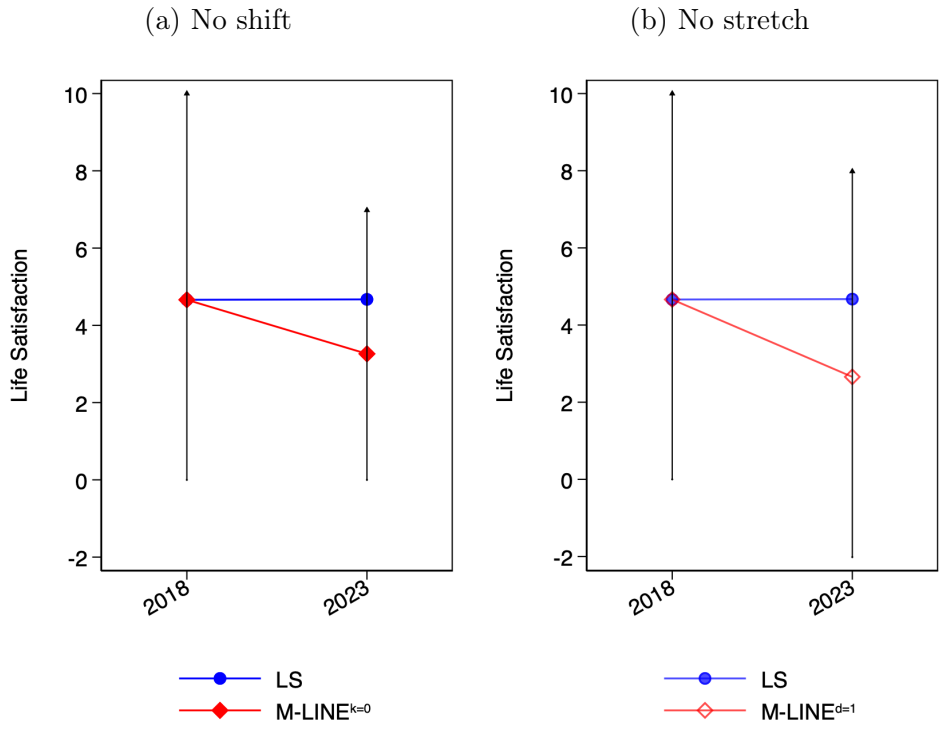


Figure 4: Wellbeing during the war: M-LINE and nominal LS in Ukraine.

Reading note: Panel (a) shows the M-LINE estimated under the no-shift hypothesis ($k = 0$). Panel (b) shows the M-LINE estimated under the no-stretch hypothesis ($d = 1$). The upward arrows represent the life satisfaction scale adjusted for rescaling. LS question: "Think about a picture of a ladder. Suppose that the top of the ladder represents the best possible life for you, and the bottom represents the worst possible life for you. If the top step is "10" and the bottom step is "0", on which step of the ladder do you feel you personally stand at the present time?" [0-10]. RLS question: "On which step would you say you stood five years ago?" [0-10]. Source: Gallup World Poll. Data is from different individuals interviewed in two nationally representative surveys of the Ukrainian population in 2018 and 2023.

(SOEP); (ii) an online panel of Prolific participants resident in the U.K. and interviewed in February 2023 and February 2024.

Our first data source of data is SOEP, i.e. the only existing large social survey panel that asked about life satisfaction and retrospective life satisfaction (until 1987). The events that we consider are: getting married, becoming widower, moving to full-time employment (from inactive/unemployed), becoming unemployed, developing a new disability, and having a kid. Table 4 reports some summary statistics on the frequency of these events in our sample.

To the purpose of testing construct validity, we estimate the following model using a first difference estimator with clustered standard errors at the individual level:

$$(h_{it} - h_{it-1}) = (X_{it} - X_{it-1})'\beta + (log(income)_{it} - log(income)_{it-1})'\gamma + (\epsilon_{it} - \epsilon_{it-1}) \tag{15}$$

where $h_{it} - h_{it-1}$ is a measure of wellbeing change (either ΔLS or ΔM-LINE), X_{it} is a set of binary variables that take value 1 if the event happened in the last year, and ϵ_{it} is a normally distributed error term. The estimation of individual-specific parameters requires two small adjustments in the calculation of d: (i) we winsorize the bottom of the life satisfaction scale at 1, rather than 0; and (ii) we trim 1% of the dataset to exclude the most extreme cases of rescaling. The adjustments are required only for the panel analyses in the present section (table 5, columns (3) and (4); table 3, column (2)).[17]

Table 3 reports the results of the regressions. The columns differ in the regressand, i.e., how changes in latent satisfaction are measured. Column (1) uses the standard approach, where differences in h are equated to differences in LS levels. Column (2) uses the M-LINE in the no-shift scenario ($k = 0$), while column (3) uses the M-LINE in the no-stretch scenario ($d = 1$).

The R^2 provide a direct measure of construct validity, by showing the correspondence between the wellbeing variable and the events that we would expect to be related to it. They show that these events explain about 1% of the variations in nominal LS levels (this is similar to Kaiser and Prati, 2024, that run a similar exercise on the British Household Panel Survey). Instead, the amount of explained variation in the M-LINE is 1.5 to 2 times higher. This is consistent with the idea that positive and negative events affect not only LS, but also the LS scale in the same direction, so that the observed change in nominal LS is attenuated.

The estimated coefficients are mostly consistent across specifications, and with the expected sign, although nominal LS seems to underestimate the positive short-term impact of marriage and employment, and to overestimate the role of income. The only important exception is childbirth. A long-standing puzzle in well-being research is about why people deliberately have kids when this

[17] The adjustments deal with two issues that can arise in the calculation of the ratio LS_{t-1}/RLS_t. The first one is that the stretching parameter is undetermined when $RLS_t = 0$ and is null whenever $LS_{t-1} = 0$. This never happens at the aggregate level (because the mean is never zero) and, although 0/10 is a rare answer at the individual level (less than 1% of respondents choose this option), the computation of the M-LINE should accommodate these cases. To get around this problem, we winsorize the bottom of the life satisfaction scale, so that all answers that are "0" on the 0-10 or 0-100 scales are counted as "1". The second issue relates to the fact that a few respondents report big differences between LS_{t-1} and RLS_t, which generates implausible values for the stretching parameter (up to a tenfold expansion of the scale), and for the shifting parameter (up to a 10-point shift in one year). We therefore apply some trimming, and discard the 0.5% of observations in the two tails of the distributions of d and k. Eventually, no more than 2% of the observations are discarded as outliers (1% due to d and 1% because of k).

is detrimental to their happiness. The M-LINE offers a different interpretation: people are happier post-kids than pre-kids, but the event can trigger substantial rescaling, up to a point where it lowers apparent nominal satisfaction.[18]

SOEP has some limitations. It allows to study only a limited set of events on a population from 40 years ago. Since it is the only panel social survey that has ever collected information on retrospective life satisfaction and life satisfaction, we run our own data collection and study the UK online panel.

In both waves of the U.K. panel, participants were asked to rate their life satisfaction on a continuous 0-100 scale. In wave 2, they were also asked to report their reconstructed life satisfaction as of one year ago, again on a 0-100 scale, as well as detailed information about events that could have happened in their lives. We adopt the exact same estimation strategy as with SOEP, and use first-difference. In the regressions, we consider two sets of events: a restrained set (table 5, columns (1), (3) and (5)) and a complete set (columns (2), (4) and (6)). The restrained set includes the following events: having a child, starting a new romantic relationship, getting fired, quitting a job, contracting a debt, finishing paying a debt, being diagnosed a new medical condition / incur a physical injury, starting a new job, getting promoted, experiencing bereavement / ceasing contact with a close one. The frequency of the events from the restricted set in our online sample is reported in table 6. The additional 17 events of the extended set are detailed in the legend of table 5.

The results are largely consistent with the SOEP analyses. The increased granularity of LS and RLS (now measured on a 100-point scale), together with the richer set of regressors, helps increasing the R squared substantially. An eye on the R squared across specifications shows that both $\Delta\text{MLINE}^{k=0}$ and $\Delta\text{MLINE}^{d=1}$ significantly outperform LS in terms of goodness of fit. Across specifications, the estimated coefficients carry the same (expected) sign in most of the recorded events. The M-LINE seems to be particularly sensitive in capturing the effect of positive financial changes, such as increasing one's own income and extinguishing a debt. The effect of childbirth, which is null when considering nominal LS, turns positive when correcting for rescaling.

Overall, the two exercises concur in showing that the M-LINE is better at post-dicting what happened in people's lives during the past year than what LS levels do.

5.2 Construct validity with *subjective* variables: does the M-LINE measure if people feel better or worse with their lives?

The M-LINE is intended to measure the change in latent life satisfaction, while correcting for rescaling. Therefore, if someone's M-LINE increases over a time period then they should *feel* that they are *more* satisfied with their life today; vice versa, if the M-LINE declines, they should *feel* that they are *less* satisfied with their life today. A straightforward way to test this prediction is to measure both the M-LINE and the answer to the question "Compared to one year ago, how satisfied with

[18] To be precise, the birth of a child can cause re-conceptualization and re-prioritization, that is, a change in the meaning and ingredients of the top of the scale, rather than in the **level** of aspirations.

Table 3: First-difference regression on life events (SOEP, 1985-87)

	(1) ΔLS	(2) ΔMLINE$^{k=0}$	(3) ΔMLINE$^{d=1}$
income change (log)	0.26***	0.16***	0.17***
	(0.04)	(0.04)	(0.03)
childbirth	-0.16***	0.21***	0.07
	(0.06)	(0.07)	(0.04)
marriage	0.20**	0.59***	0.39***
	(0.10)	(0.11)	(0.07)
start employment	0.25***	0.78***	0.39***
	(0.06)	(0.08)	(0.05)
start unemployment	-0.87***	-0.53***	-0.72***
	(0.09)	(0.09)	(0.07)
new disability	-0.13	0.00	-0.16**
	(0.12)	(0.12)	(0.07)
bereavement	-0.64**	-0.41	-0.89***
	(0.29)	(0.37)	(0.29)
N	27,345	27,345	27,345
R^2	0.010	0.016	0.020

Reading note: Linear regressions of change individual changes in life satisfaction on life events. Column (1) uses nominal life satisfaction; columns (2) uses the M-LINE under the no-shift scenario; columns (3) uses the M-LINE under the no-stretch scenario. Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. LS question: "How satisfied are you at present with your life as a whole?" [0-10]. RLS question: "How satisfied were you a year ago with your life?" [0-10].

your life are you nowadays?", measured on a continuous scale from -100 to +100.[19] We label it the "feeling better/worse" variable. The estimated correlation of this variable with the M-LINE provides a direct test of convergent validity.

We collected the necessary information from the UK online panel described above (N=987). Table 7 displays the correlation matrix of ΔLS, ΔM-LINE (no-stretch and no-shift scenarios) and the "feeling better/worse" variable. The estimated Pearson correlation between ΔLS and ΔM-LINE is +39%/+49%, reflecting the idea that LS and M-LINE capture similar yet not identical information. The estimated correlation between ΔLS and the direct question about "feeling better/worse" is also positive (+34%), as expected. Crucially, the M-LINE has even stronger correlations with people's report of "feeling better/worse". The estimated correlation is +0.53% in the no-shift scenario and +0.60% in the no-stretch scenario.

Overall, the M-LINE seems to capture not only what is going on in people's life (as witnessed by the correlation with objective variables), but also what is on people's mind (as shown by the correlation with the subjective feeling of doing better/worse).

[19]Recently, Kaiser and Prati (mimeo) studied the psychometric properties of comparative life satisfaction ratings and concluded that they outperform nominal life satisfaction levels on a variety of criteria.

Table 4: Recorded Events (SOEP, 1984-87)

Category	Nb of occurrences	N
Marriage	432	27,345
Bereavement	69	27,345
Employment	1,257	27,345
Unemployment	674	27,345
New disability	345	27,345
Childbirth	934	27,345
Income change*	24,237	27,345

* There were 24,237 episodes of annual household income changes. Income change is coded as Δ log(income). On average, the annual change was about +3%.

5.3 *Objective* reliability: Test-retest reliability of RLS

If RLS is a reliable measure of past latent satisfaction, then the reconstructed satisfaction of a particular period should not change except for rescaling effects. We test this hypothesis using the US online panel (N=1,647). The RLS score should be the same when measured at two time points that are too close to expect any rescaling. Therefore, we compare two recall questions that refer to the same period but were asked at two different points of time, with a two-week interval (January 3, and January 17-20, 2024). On both dates, respondents were asked to evaluate their life 5 years ago, using the Cantril ladder. On average, they gave the same response in the two questionnaires ($RLS_{t,t-5y} = RLS_{t-2w,t-5y}$; $diff = 0.04$; $p - value = 0.208$]). This should contribute to rule out one of the main suspicions about reconstructed satisfaction, i.e. some low reliability.

5.4 *Subjective* reliability: Perceived accuracy of RLS

One could wonder whether remembered satisfaction can be reliably used or whether it merely captures the noise of memory errors. People struggle to recall their past satisfaction and may be uncertain about their response. If this is the case, RLS could provide quite noisy information about h_{t-1}. To estimate the perceived accuracy of retrospective satisfaction, we rely on the concept of "cognitive uncertainty" developed by Enke and Graeber (2023). In our UK online panel, after asking participants to report their life satisfaction one year ago on a continuous scale, we asked them: "How certain are you that your life satisfaction was exactly $x/100$?" where x corresponds to RLS. Participants would then report a lower- and upper-bound for RLS.[20] We adopted the same procedure to estimate the amount of cognitive uncertainty about current life satisfaction, and compared the two.

Table 9 summarizes the results. The average individual in the sample is *sure* that their current life satisfaction, on a scale from 0 to 100, is between 56 and 73, and that their life satisfaction one year ago was between 54 and 70. The difference between the upper- and lower-bound provides a

[20]The response was formulated as follows; "I am certain that my life satisfaction was actually:
At the very least:
At the very most: ".

Table 5: First-difference regression on life events (UK, 2023-24)

	(1) ΔLS	(2) ΔLS	(3) ΔMLINE$^{k=0}$	(4) ΔMLINE$^{k=0}$	(5) ΔMLINE$^{d=1}$	(6) ΔMLINE$^{d=1}$
income change (log)	4.31**	4.58**	11.30***	11.18***	8.40***	8.21***
	(1.98)	(2.04)	(2.91)	(2.94)	(1.93)	(1.89)
childbirth	3.04	2.97	5.44**	4.53*	5.12***	4.36**
	(2.97)	(3.01)	(2.62)	(2.62)	(1.78)	(1.85)
new partner	2.22	4.51	7.58**	8.20**	3.97	4.57
	(2.61)	(2.76)	(3.33)	(3.28)	(3.25)	(3.31)
fired	-2.91	-1.82	1.06	0.43	-2.19	-2.17
	(2.99)	(3.27)	(5.19)	(5.19)	(3.17)	(3.24)
quit job	-0.39	0.12	1.99	1.51	2.48	2.36
	(2.39)	(2.40)	(3.08)	(3.14)	(2.34)	(2.40)
start debt	-3.35**	-2.66*	-0.99	-1.78	-1.89	-2.27*
	(1.58)	(1.58)	(1.55)	(1.58)	(1.32)	(1.35)
end debt	0.88	0.96	4.05***	3.40***	3.42***	2.83***
	(1.12)	(1.16)	(1.23)	(1.29)	(1.01)	(1.04)
new diagnosis/injury	-1.73	-0.11	-0.83	-0.76	-1.04	-0.92
	(1.15)	(1.33)	(1.07)	(1.19)	(0.94)	(1.07)
new job	3.23*	3.31*	5.95**	4.93*	3.35**	2.73*
	(1.76)	(1.75)	(2.93)	(2.76)	(1.66)	(1.60)
promoted	-0.21	-0.19	2.68	2.12	3.14**	2.90**
	(1.51)	(1.57)	(1.67)	(1.78)	(1.37)	(1.44)
bereavement/loss	-1.74	-1.53	0.63	-0.19	-0.37	-0.77
	(1.09)	(1.23)	(1.02)	(1.15)	(0.90)	(0.94)
Additional events		✓		✓		✓
N	987	987	987	987	987	987
R^2	0.029	0.061	0.111	0.139	0.100	0.124

Reading note: Linear regressions of change individual changes in life satisfaction on life events. Columns (1)-(2) use nominal life satisfaction; columns (3)-(4) use the M-LINE under the no-shift scenario; columns (5)-(6) use the M-LINE under the no-stretch scenario. Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. LS question: "How satisfied are you with your life, all things considered?" [0-100]. RLS question: "And how satisfied were you with your life about one year ago, all things considered?" [0-100]. Specifications (2), (4) and (6) include the following additional binary covariates (= 1 if experienced by the respondent in the last 12 months): Hospitalized; Started therapy; Recovered from illness; Retired; Bought a house; Moved dwelling; New person in household; New pet; New close person; Took a long holiday; Separated from partner; A close person got sick/injured; Other health event; Other job event; Other housing event; Other financial event; Other romantic event; Other personal event.

Table 6: Recorded events (UK, 2023-24)

Category	**Nb of occurrences**	**N**
Childbirth	28	987
New Partner	36	987
Fired	24	987
Quit Job	67	987
Start Debt	133	987
End Debt	180	987
New Diagnosis/Injury	169	987
New Job	136	987
Promoted	89	987
Bereavement/Loss	258	987
income change*	648	987

* There were 658 episodes of annual personal income changes. Income change is coded as Δlog(income). On average, the annual change was about +7%.

Table 7: Correlation Matrix

	ΔLS	ΔM-LINE$^{d=1}$	ΔM-LINE$^{k=0}$	Feeling better/worse
ΔLS	1.0000			
ΔM-LINE$^{d=1}$	0.49*	1.0000		
ΔM-LINE$^{k=0}$	0.39*	0.89*	1.0000	
Feeling better/worse	0.34*	0.60*	0.53*	1.0000

Table 8: RLS at different points in time

	In week 1	In week 3	diff.	p-value	N
$RLS_{2024,2019}$	5.47	5.43	0.04	0.208	1,647

measure of cognitive uncertainty, which is very similar in both cases (around 16%-17%). This result shows that respondents feel fairly accurate in how they recall their past satisfaction, to a comparable extent as to how accurate they feel in reporting their current satisfaction with life.

Table 9: RLS at different points in time

	Lower bound	Upper bound	Cognitive uncertainty	N
LS	56.5	73.5	17.0	913
RLS	53.9	69.9	16.0	913

Reading note: The sample size is smaller because some responses failed the consistency check. Typically, they reported a lower-bound that is higher than the upper-bound, of vice-versa.

5.5 Validation of the no-stretch and no-scale hypotheses: Subjective perceptions of rescaling

Is rescaling an abstract "ivory tower" possibility, or is it something people actually perceive? To inquire, we included some questions in the U.S. online panel survey (N=1,647). We asked participants whether they perceive a change in the world of possibilities over the past five years, and in which direction. The two questions were (i) "How has your idea of the 'best possible life' changed in different aspects compared to five years ago? Would you say that your idea of the 'best possible life', rated at 10/10, is higher, lower, or about the same as it was five years ago? For instance, if five years ago someone considered that an annual income of 200,000 USD was 10/10 for personal income, but today they consider that 250,000 USD is 10/10, then their best possible income is higher today than five years ago."; and (ii) "Similarly, how has your idea of the 'worst possible life' changed in different aspects compared to five years ago? Would you say that your idea of the 'worst possible life", rated 0/10, is higher, lower or about the same as it was five years ago?".

It turns out that four out of ten (41%) felt that the top of the scale increased over the last five years, and just as many (38%) felt that it did not change, while the remaining (20%) perceived a decrease. Regarding the other end of the scale, half of the participants responded that the bottom of the scale had not changed, while 31% perceived it as lower today and 17% as higher (1% N/A) than 5 years ago. This information constitutes a piece of suggestive evidence that rescaling, in particular at the top of the scale, is not only theoretically plausible, but also subjectively perceived.

5.6 You want to correct for rescaling, now what? Some recommendations

At this point, we hope the readers are convinced of the usefulness of correcting for rescaling. To implement the M-LINE from a panel of life satisfaction data, two additional pieces of information are needed: (i) reconstructed life satisfaction; (ii) the memory bias parameter.

Reconstructed life satisfaction can be measured by adding a question to the survey. Ideally, the time lag (5 years ago, 1 year ago, etc.) should correspond to the frequency of the data collection. For example, an annual panel should ask about life satisfaction one year ago. This is the best solution also in a repeated cross-section.

The memory bias parameter can be estimated either from a complementary data collection (as we did in this paper) or from the main survey. To implement the latter solution, which is probably the most practical, data producers can ask about retrospective life satisfaction two weeks ago and take the delta between the average answer to that question and current life satisfaction as a proxy for memory bias. Indeed, insofar as the latent satisfaction and the satisfaction scale can be assumed to be constant (on average) over a two-week horizon, the way respondents view their satisfaction two weeks ago reveals the average memory bias. Instead, we do *not* recommend using the $\hat{m} = 0.12$ for non-Western countries, since the memory bias could be sensitive to cultural and demographic factors that cannot be plausibly assumed constant with respect to our calibration study.

Finally, we remind that, in micro-regressions, data practitioners must use a scale with a lower bound at 1 (*not* 0) and we recommend using some trimming that discards implausible rescaling. For more details, we invite readers to review section 5.1.

6 Conclusion

We have explored the consequences of interpreting the two Easterlin paradoxes in terms of rescaling rather than hedonic treamdill. Our method is based on the exploitation of current and retrospective self-reported scores of life satisfaction. Our main assumption is that if a person, at a given point in time, declares a higher score of current life satisfaction and a lower score of past satisfaction, this gap contains some truth, and is not a pure recall bias. Accordingly, we interpret the difference between the level of past satisfaction as declared in two different points of time as partly due to a rescaling phenomenon, i.e. a shift or a stretch (or a shrink) of the latent life satisfaction scale.

We follow the rescaling conjecture and use current and retrospective self-reported levels of life satisfaction from unexplored historical data in order to reconstruct an index of latent happiness over the long run, the M-LINE. The M-LINE shows that latent life satisfaction of the average American increased substantially, from 1959 to 2008, at a pace comparable to objective measures of income or human development. Conversely, national happiness contracted in France and Germany during the COVID-19 pandemic and in Ukraine during the war. Using short-run micro-panel data, we also show that the reconstructed latent life satisfaction correlates with individuals' life events better than nominal life satisfaction does. If it is true, as Kahneman and Krueger (2006, p.18) suggest, that "the invariance of reported satisfaction to material living standards and major changes in life circumstances is a reason why some economists are skeptical of the validity of measures of subjective well-being", then our study helps to make progress on this front.

We now have two happiness measures available, nominal life satisfaction and the M-LINE, which offer potentially different pictures of how well-being has evolved over time. Which should be trusted? Nominal satisfaction assumes that rescaling does not matter, while the M-LINE assumes that no

adaptation takes place. These different theoretical assumptions lead to very different practical conclusions for policymakers and statistical institutes. If we use nominal satisfaction, there is little hope to raise national happiness in a durable way (and little risk to reduce it). If the M-LINE is used instead, national happiness will be more sensitive to what happens in a society, like changes in income, health, education, pandemics, and wars.

We do not claim to have demonstrated that rescaling rather than hedonic treadmill is the actual process that explains the relative inertia of reported happiness over long episodes of human development. Rather, the ambition of this paper is to explore the conjecture of rescaling with a new method and data and to show the consequences of this assumption. In the absence of a direct measure of latent happiness, which is unobservable, at least by social scientists, we believe that excluding one interpretation or the other is simply out of reach. We also believe that there is little reason to consider the two phenomena mutually exclusive. As noted by Stevens (1958, p.633) in an ante-litteram discussion of the topic: "Sensory adaptation is a real and basic phenomenon, as we can readily testify when we walk from a sunlit street into a darkened theater. The relativity of the meanings we attach to such quantifiers as large, medium, and small is an equally genuine phenomenon".

Nominal life satisfaction and the M-LINE are based on implausibly extreme assumptions. This is why we should probably not conclude that national happiness in the US more than doubled in the second half of the 20th century. Yet, for a similar reason, we should probably not trust that national happiness in the US has also been flat. The truth is likely to be in the middle, with both rescaling and adaptation playing some role. With the two processes in place, latent life satisfaction is strictly increasing *up to* the M-LINE. This conservative view is sufficient for the Easterlin paradox, i.e. the ink bottle of the literature on subjective wellbeing, to disappear.

Our exercise has some limitations. First, we are able to reconstruct an index of *average* national happiness over time, not individual time series. This is because to date, only very limited individual-level *panel* data with current and retrospective reports of life satisfaction are available. Were it the case, we could address the initial Easterlin's question about "increasing the happiness *of all*", instead of "increasing *aggregate* happiness" as we do here. We could, in particular, analyze the influence of a person's relative position in the social stratum on her latent happiness. Our agenda for future research is to realize this exercise as soon as adequate individual-level panel data are available. Another limitation of this study is that we have one too many degrees of freedom in the identification of the rescaling process, and we run several scenarios based on different assumptions about the shift and stretch modes. No correction is assumption-free, and future research will have to establish which of the scenarios we consider is the most psychologically sound.

In spite of these drawbacks, we think that the method consisting in using simultaneously current and retrospective reports of life satisfaction could be extended to improve the measurement of time variations in subjective well-being, not only over long time spans, but also for the evaluation of public policy interventions.

References

Adler, M. D. (2019). *Measuring social welfare: An introduction*. Oxford University Press, USA.

Anderson, J. R., Bothell, D., Lebiere, C., and Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language*, 38(4):341–380.

Andreoni, J., Bernheim, B. D., and Jia, T. (2024). Do people report happiness accurately? Technical report, National Bureau of Economic Research.

Benjamin, D. J., Cooper, K., Heffetz, O., and Kimball, M. S. (2023a). From happiness data to economic conclusions.

Benjamin, D. J., Cooper, K., Heffetz, O., Kimball, M. S., and Zhou, J. (2023b). Adjusting for scale-use heterogeneity in self-reported well-being. Technical report, National Bureau of Economic Research.

Bond, T. N. and Lang, K. (2019). The sad truth about happiness scales. *Journal of Political Economy*, 127(4):1629–1640.

Brickman, P., Coates, D., and Janoff-Bulman, R. (1978). Lottery winners and accident victims: Is happiness relative? *Journal of Personality and Social Psychology*, 36(8):917.

Busseri, M. A. and Samani, M. N. (2019). Lay theories for life satisfaction and the belief that life gets better and better. *Journal of Happiness Studies*, 20:1647–1672.

Cantril, H. (1965). *Pattern of human concerns*. Rutgers University Press.

Coupe, T. and Obrizan, M. (2016). The impact of war on happiness: The case of ukraine. *Journal of Economic Behavior & Organization*, 132:228–242.

Daly, M. and Robinson, E. (2022). Depression and anxiety during covid-19. *The Lancet*, 399(10324):518.

Deaton, A. (2008). Income, health, and well-being around the world: Evidence from the gallup world poll. *Journal of Economic perspectives*, 22(2):53–72.

Diener, E. and Oishi, S. (2000). Money and happiness: Income and subjective well-being across nations.

Easterlin, R. A. (1995). Will raising the incomes of all increase the happiness of all? *Journal of economic behavior & organization*, 27(1):35–47.

Easterlin, R. A. (2001). Income and happiness: Towards a unified theory. *The Economic Journal*, 111(473):465–484.

Easterlin, R. A. (2003). Explaining happiness. *Proceedings of the National Academy of Sciences*, 100(19):11176–11183.

Easterlin, R. A. and O'Connor, K. J. (2023). Three years of covid-19 and life satisfaction in europe: A macro view. *Proceedings of the National Academy of Sciences*, 120(19):e2300717120.

Enke, B. and Graeber, T. (2023). Cognitive uncertainty. *The Quarterly Journal of Economics*, 138(4):2021–2067.

Fabian, M. (2022a). Scale norming undermines the use of life satisfaction scale data for welfare analysis. *Journal of Happiness Studies*, 23(4):1509–1541.

Fabian, M. (2022b). *A theory of subjective wellbeing*. Oxford University Press.

Frederick, S. and Loewenstein, G. (1999). Hedonic adaptation. In Kahneman, D. Schwarz, N., editor, *Well-Being the Foundations of Hedonic Psychology*.

Hagerty, M. R. (2003). Was life better in the "good old days"? intertemporal judgments of life satisfaction. *Journal of Happiness Studies*, 4:115–139.

Helliwell, J. F., Huang, H., Wang, S., and Norton, M. (2021). World happiness, trust and deaths under covid-19. *World happiness report*, 2021:13–57.

Helliwell, J. F., Layard, R., Sachs, J. D., and Neve, J.-E. D. (2022). World happiness report 2022.

HM Treasury (2021). Wellbeing guidance for appraisal: Supplementary green book guidance. Technical report.

Howard, G. S. and Dailey, P. R. (1979). Response-shift bias: A source of contamination of self-report measures. *Journal of Applied Psychology*, 64(2):144.

Howard, G. S., Dailey, P. R., and Gulanick, N. A. (1979a). The feasibility of informed pretests in attenuating response-shift bias. *Applied Psychological Measurement*, 3(4):481–494.

Howard, G. S., Schmeck, R. R., and Bray, J. H. (1979b). Internal invalidity in studies employing self-report instruments: A suggested remedy. *Journal of Educational Measurement*, pages 129–135.

IMF (2024). *October 2024 World Economic Outlook*.

Kahana, M. J. (2012). *Foundations of human memory*. OUP USA.

Kahneman, D. (2009). The sad tale of the aspiration treadmill. In Brockman, J., editor, *What Have You Changed Your Mind About?: Today's Leading Minds Rethink Everything*. Edge.

Kahneman, D. and Krueger, A. B. (2006). Developments in the measurement of subjective well-being. *Journal of Economic Perspectives*, 20(1):3–24.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):363–391.

Kaiser, C. (2022). Using memories to assess the intrapersonal comparability of wellbeing reports. *Journal of Economic Behavior & Organization*, 193:410–442.

Kaiser, C. and Oswald, A. J. (2022a). Inequality, well-being, and the problem of the unknown reporting function. *Proceedings of the National Academy of Sciences*, 119(50):e2217750119.

Kaiser, C. and Oswald, A. J. (2022b). The scientific value of numerical measures of human feelings. *Proceedings of the National Academy of Sciences*, 119(42):e2210412119.

Kaiser, C. and Prati, A. (2024). Why you should measure subjective changes. *Unpublished manuscript*.

Kaiser, C. and Vendrik, M. (2020). How threatening are transformations of happiness scales to subjective wellbeing research?

Kapteyn, A., James, S., and Arthur, v. S. (2007). Vignettes and Self-reports of Work Disability in the U.S. and the Netherlands. *American Economic Review*, 97(1):946–975.

King, G., Christopher, M., Joshua, S., and Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, 98(1):191–207.

Klemm, M. (2022). Well-being changes from year to year: A comparison of current, remembered and predicted life satisfaction. *Journal of Happiness Studies*, 23(4):1669–1681.

Köke, S. and Perino, G. (2017). For "better" or "worse": a direct approach to elicit preference rankings from life-satisfaction data. *WiSo-HH Working Paper Series*, (43).

Krueger, A. B. and Schkade, D. A. (2008). The reliability of subjective well-being measures. *Journal of public economics*, 92(8-9):1833–1845.

Lykken, D. and Tellegen, A. (1996). Happiness is a stochastic phenomenon. *Psychological Science*, 7(3):186–189.

Mahoney, J. (2023). Subjective well-being measurement: Current practice and new frontiers.

Nilsson, A. H., Eichstaedt, J. C., Lomas, T., Schwartz, A., and Kjell, O. (2024). The cantril ladder elicits thoughts about power and wealth. *Scientific Reports*, 14(1):2642.

Odermatt, R. and Stutzer, A. (2019). (mis-) predicted subjective well-being following life events. *Journal of the European Economic Association*, 17(1):245–283.

Oort, F. J., Visser, M. R., and Sprangers, M. A. (2005). An application of structural equation modeling to detect response shifts and true change in quality of life data from cancer patients undergoing invasive surgery. *Quality of life Research*, 14:599–609.

Oswald, A. J. (2008). On the curvature of the reporting function from objective reality to subjective feelings. *Economics Letters*, 100(3):369–372.

Oswald, A. J. and Powdthavee, N. (2008). Does happiness adapt? A longitudinal study of disability with implications for economists and judges. *Journal of public economics*, 92(5-6):1061–1077.

Palan, S. and Schitter, C. (2018). Prolific.ac−A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17:22–27.

Prados de la Escosura, L. (2021). Augmented human development in the age of globalization. *The Economic History Review*, 74(4):946–975.

Prati, A. and Senik, C. (2022). Feeling good is feeling better. *Psychological Science*, 33(11):1828–1841.

Schwartz, C. E. and Sprangers, M. A. (2000). Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research.

Schwartz, C. E., Stucky, B., Rivers, C. S., Noonan, V. K., Finkelstein, J. A., and Network, R. (2018). Quality of life and adaptation in people with spinal cord injury: response shift effects from 1 to 5 years postinjury. *Archives of Physical Medicine and Rehabilitation*, 99(8):1599–1608.

Socio-Economic Panel (2023). Data from 1984-2021 (SOEP-Core, v38, Remote Edition). `doi: 10.5684/soep.core.v38r`.

Stevens, S. S. (1958). Adaptation-level vs. the relativity of judgment. *The American Journal of Psychology*, 71(4):633–646.

Stevenson, B. and Wolfers, J. (2008). Economic growth and subjective well-being: Reassessing the Easterlin paradox. *National Bureau of Economic Research*.

Stillman, S., Gibson, J., McKenzie, D., and Rohorua, H. (2015). Miserable migrants? natural experiment evidence on international migration and objective and subjective well-being. *World Development*, 65:79–93.

Stutzer, A. (2004). The role of income aspirations in individual happiness. *Journal of Economic Behavior & Organization*, 54(1):89–109.

Van Leeuwen, C. M., Post, M. W., Van Der Woude, L. H., De Groot, S., Smit, C., Van Kuppevelt, D., and Lindeman, E. (2012). Changes in life satisfaction in persons with spinal cord injury during and after inpatient rehabilitation: adaptation or measurement bias? *Quality of Life Research*, 21:1499–1508.

Van Praag, B. (1971). The welfare function of income in Belgium: An empirical investigation. *European Economic Review*, 2(3):337–369.

Van Praag, B. and Frijters, P. (1999). The measurement of welfare and well-being: The Leyden approach. In Diener, E., Kahneman, D., and Schwarz, N., editors, *Well-being: The foundations of hedonic psychology*, pages 413–433. Russel Sage Foundation New York.

Van Praag, B. M. (1991). Ordinal and cardinal utility: an integration of the two dimensions of the welfare concept. *Journal of econometrics*, 50(1-2):69–89.

Van Praag, B. M., Dubnoff, S., and Van Der Sar, N. L. (1988). On the measurement and explana-

tion of standards with respect to income, age and education. *Journal of Economic Psychology*, 9(4):481–498.

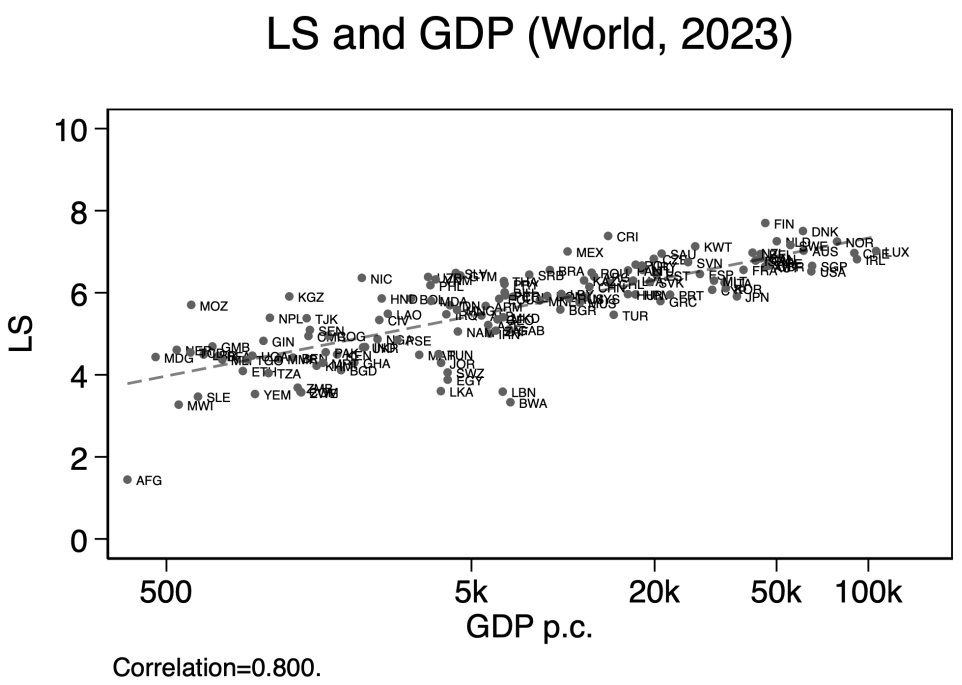
Yin, R., Lepinteur, A., Clark, A. E., and D'Ambrosio, C. (2023). Life satisfaction and the human development index across the world. *Journal of Cross-Cultural Psychology*, 54(2):269–282.

Acknowledgments

We are grateful for financial support from the French National Research Agency, through the Investissements d'avenir framework ANR-17-EURE-0001; from the AAPG2021(WRKCOV19); from CEPREMAP; from the John Fell Fund; and from the Oxford Wellbeing Research Centre. We benefited from the remarks and suggestions of participants in seminars in Bordeaux University, Frei University (Berlin), Luxembourg University, the London School of Economics, the Paris School of Economics, Keio University, University College London, and at the Conference of European Statistics Stakeholders, as well as discussion with Farid Anvari, Roland Bénabou, Andrew Clark, Jan-Emmanuel De Neve, Mark Fabian, Marc Fleurbaey, Caspar Kaiser, Carol Graham, Andrew Oswald, Nicolas Jacquemet, Leonardo Pejsachowicz, Leandro Prados de la Escosura, Marco Ranaldi. We would like to thank the Roper Centre for its support with the historical US records and Rachel Penrod for her support with the more recent U.S. records, as well as Gallup.Inc and Wellbeing for Planet Earth for making the global data collection possible. Rémy Bellaunay and Isaac Parkes provided excellent research assistance. All errors are our own.

Appendices

A1 Additional Figures



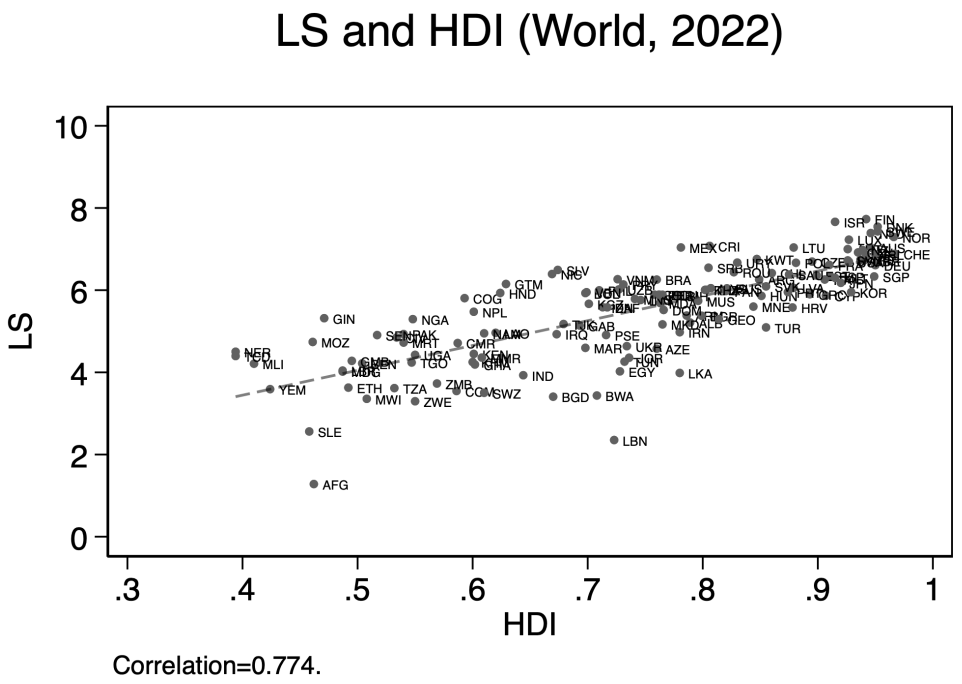
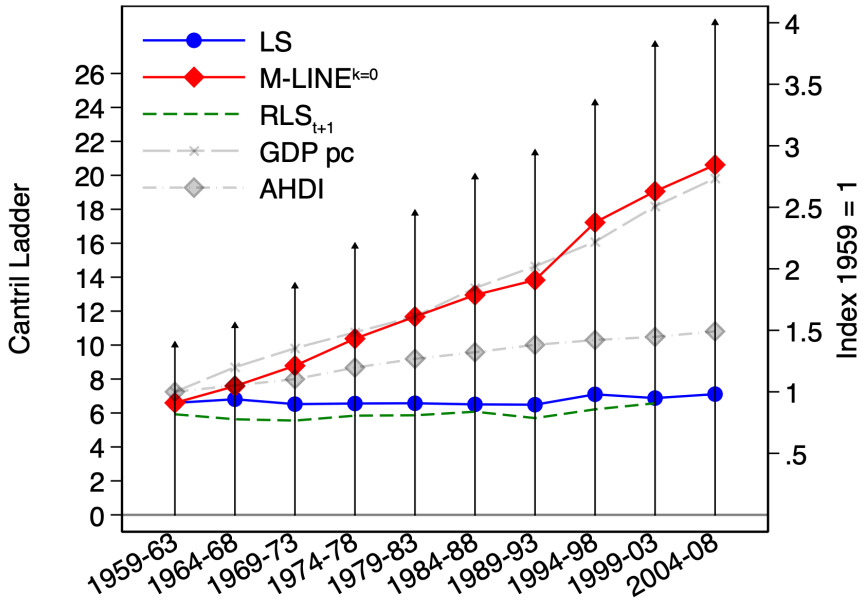
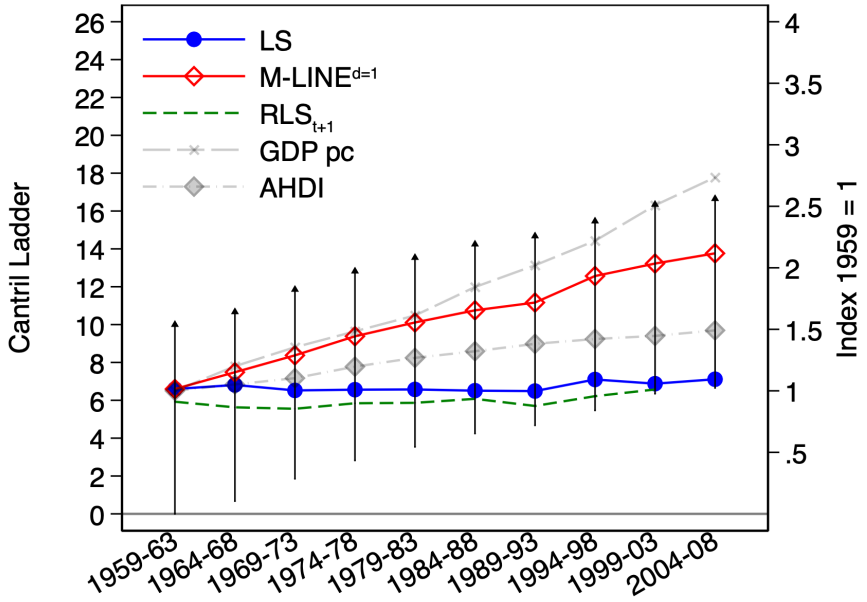


Figure A1: LS and objective wellbeing (cross-section).

At a point in time, richer and more developed countries are also happier. $N = 133$ countries. The y-axis indicates life-satisfaction scale points, as measured by the Cantril Ladder. GDP p.c. is expressed on a logarithmic scale and is imputed with the value of the previous year whether missing. Data sources: Gallup World Poll; World Bank; United Nations Development Programme.

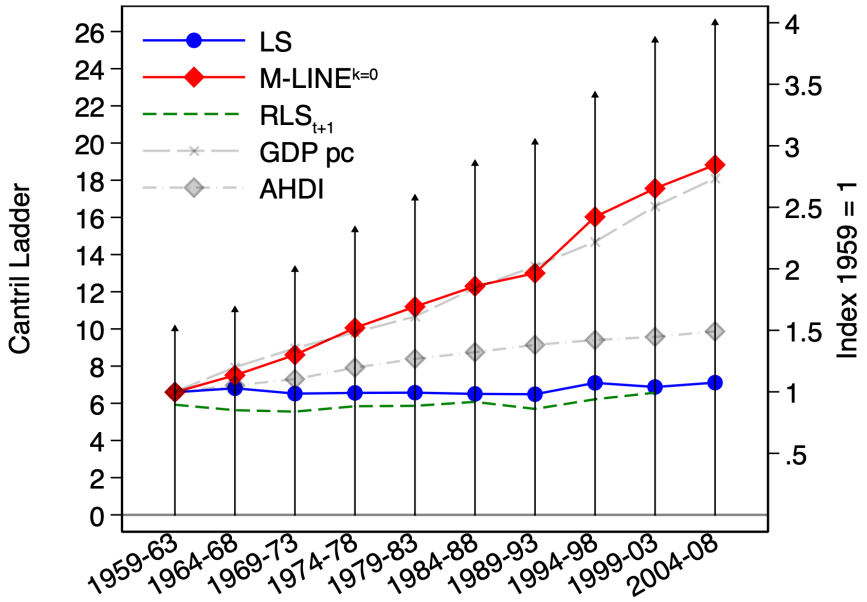


(a) No shift (m=**0.00**)

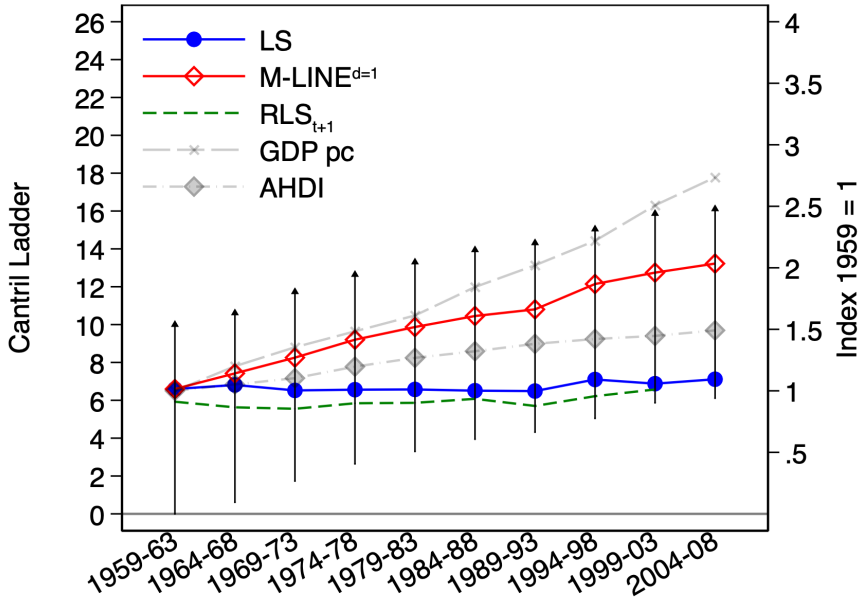


(b) No stretch (m=**0.00**)

Figure A2: Latent satisfaction for different rescaling patterns (USA, 1959-2008) for m=**0.00**

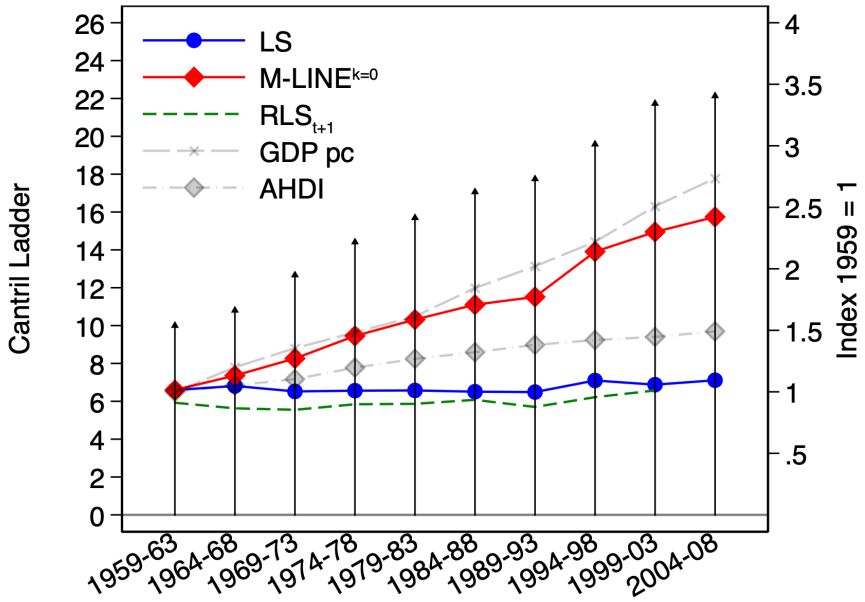


(a) No shift (m=**0.06**)

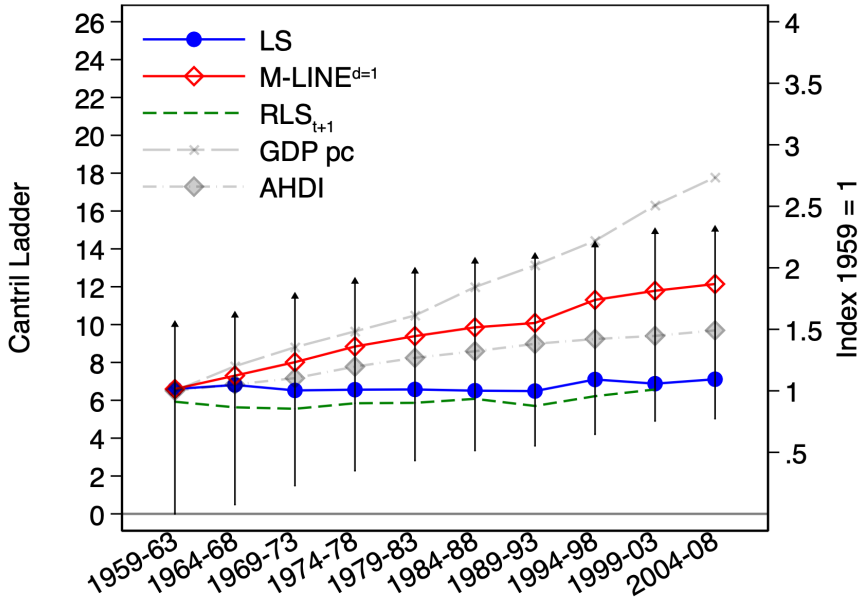


(b) No stretch (m=**0.06**)

Figure A3: Latent satisfaction for different rescaling patterns (USA, 1959-2008) for m=**0.06**

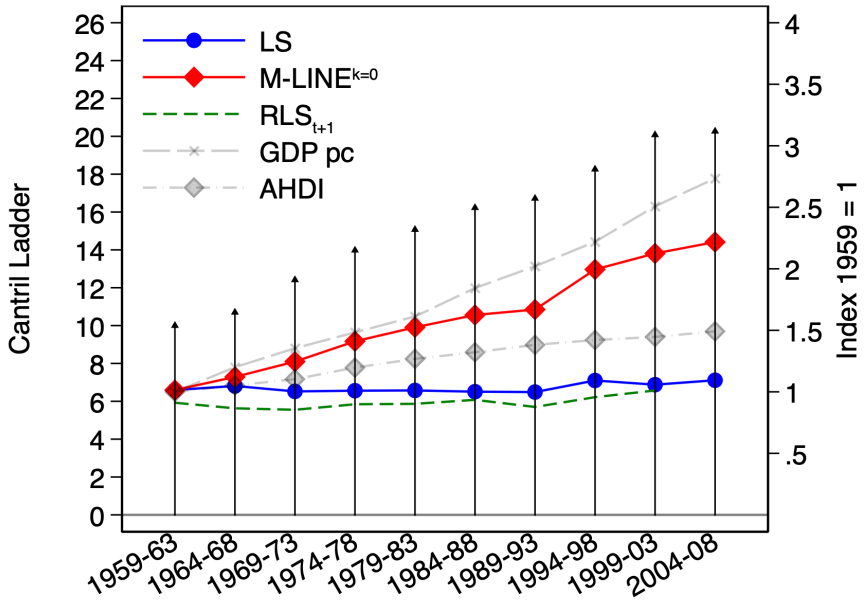


(a) No shift (m=**0.18**)

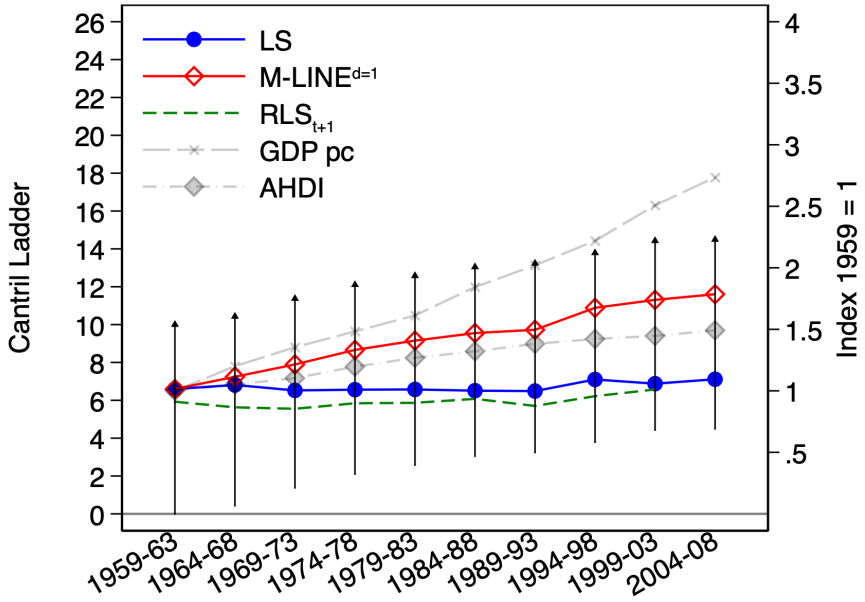


(b) No stretch (m=**0.18**)

Figure A4: Latent satisfaction for different rescaling patterns (USA, 1959-2008) for m=**0.18**

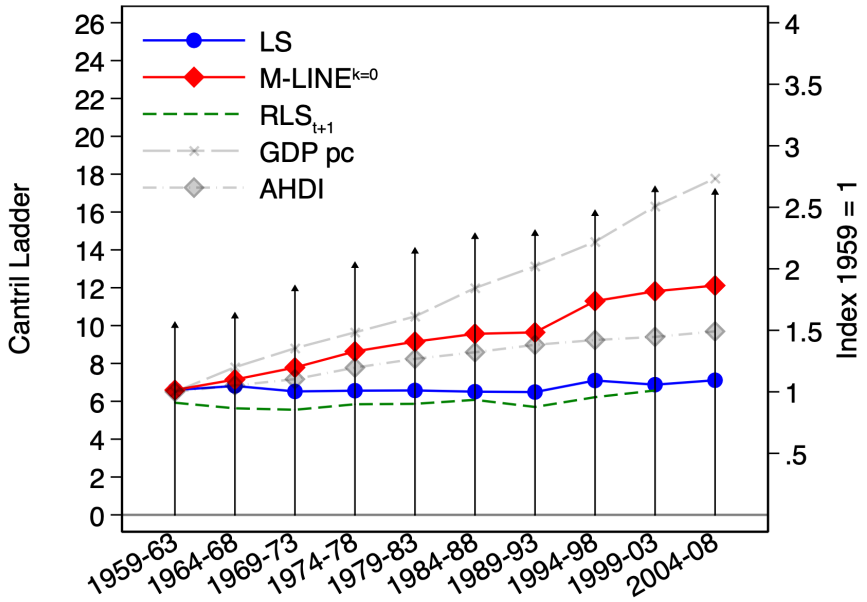


(a) No shift (m=**0.24**)

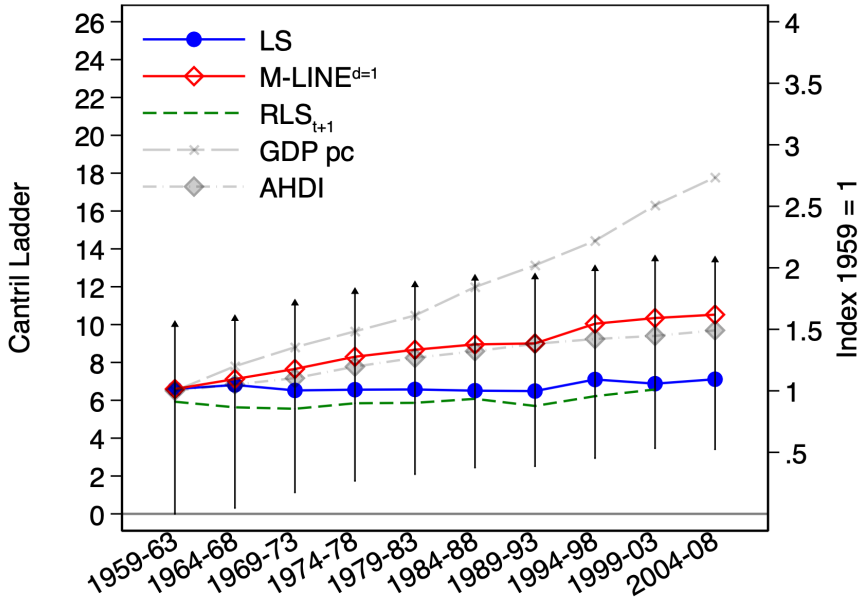


(b) No stretch (m=**0.24**)

Figure A5: Latent satisfaction for different rescaling patterns (USA, 1959-2008) for m=**0.24**

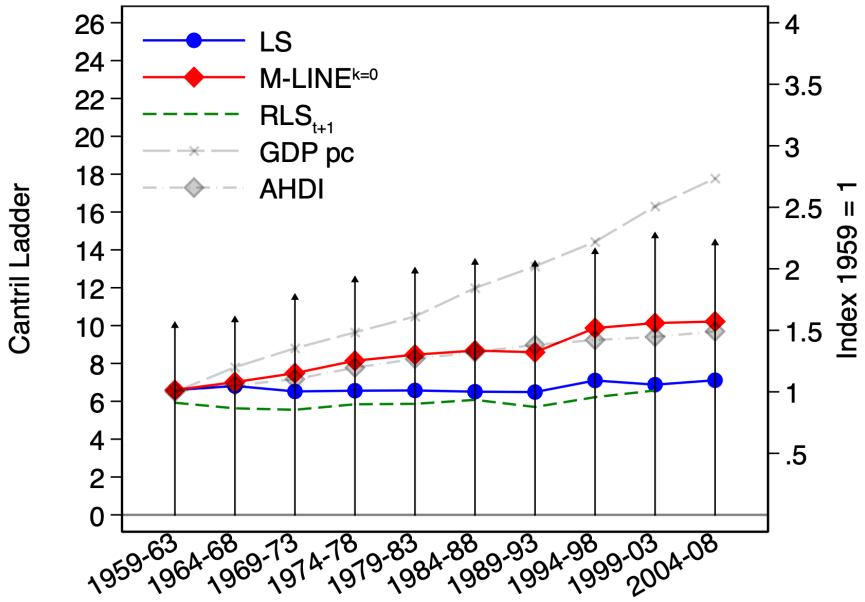


(a) No shift (m=**0.36**)

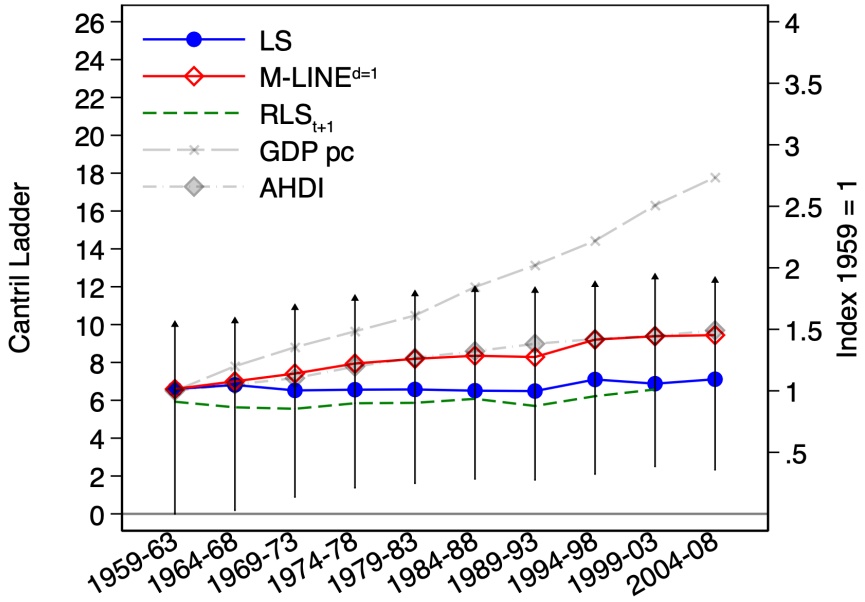


(b) No stretch (m=**0.36**)

Figure A6: Latent satisfaction for different rescaling patterns (USA, 1959-2008) for m=**0.36**

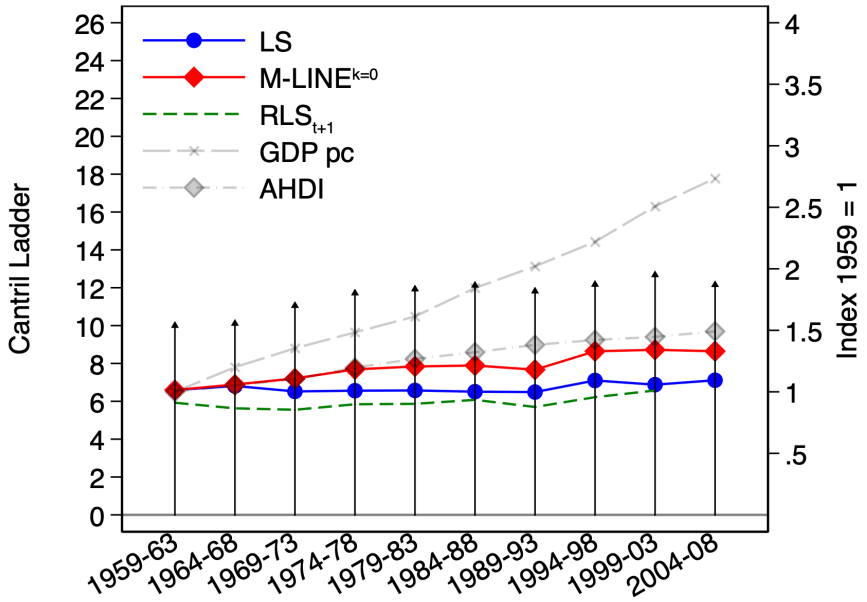


(a) No shift (m=**0.48**)

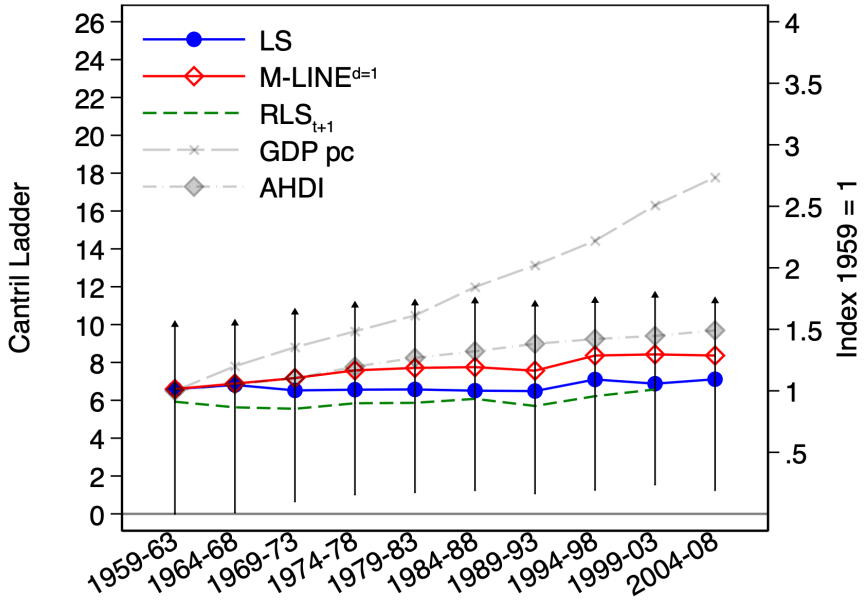


(b) No stretch (m=**0.48**)

Figure A7: Latent satisfaction for different rescaling patterns (USA, 1959-2008) for m=**0.48**



(a) No shift (m=**0.60**)



(b) No stretch (m=**0.60**)

Figure A8: Latent satisfaction for different rescaling patterns (USA, 1959-2008) for m=**0.60**

A2 Additional Tables

	1959-63	1964-68	1969-1973	1974-1978	1979-1983
GDP pc	18,562	22,280	24,931	27,540	29,939
Life exp	69.93	70.26	70.89	72.52	73.92
Edu Att	45.00	50.12	56.03	63.72	69.82
Dem Index	0.55	0.58	0.64	0.73	0.78
RLS	5.91	5.92	5.62	5.55	5.84
LS	6.59	6.81	6.52	6.56	6.57
			No shift		
$\hat{m}^*$	0.12	0.12	0.12	0.12	0.12
k^*	0.00	0.00	0.00	0.00	0.00
d	1.00	1.09	1.29	1.49	1.63
h_{min}	0.00	0.00	0.00	0.00	0.00
h_{max}	10.00	10.91	12.93	14.86	16.35
M-LINE	6.59	7.43	8.43	9.75	10.75
			No stretch		
$\hat{m}^*$	0.12	0.12	0.12	0.12	0.12
k	0.00	0.55	1.61	2.46	3.06
d^*	1.00	1.00	1.00	1.00	1.00
h_{min}	0.00	0.55	1.61	2.46	3.06
h_{max}	10.00	10.55	11.61	12.46	13.06
MLINE	6.59	7.36	8.13	9.02	9.63

Table A1: Economic and Happiness Indicators for the USA on the long term

Reading note: RLS and LS are the sample mean from the values reported in the survey. $\hat{m}^*$ is a fixed parameter which was calibrated on the US online survey. k and d are, respectively, the shifting and stretching parameters. The asterisk * is used whether the parameter is fixed instead of estimated. h_{min} and h_{max} indicate the location of, respectively, the bottom and the top of the happiness scale, expressed with respect to 1959. M-LINE$^{k\,=\,0}$ and M-LINE$^{d\,=\,1}$ denote latent satisfaction after the correction for rescaling and under the hypothesis that, respectively, $k = 0$ (no shift) or $d = 1$ (no stretch). Real gross domestic product per capita is measured as chained 2012 Dollars (Source: St.Louis Federal reserve Bank). Life at expectancy at birth is measured in life years (Source: United Nations - World Population Prospects); Educational Attainment at high school is measured in percentage of the population (Source: US Census Bureau). Liberal democracy index is measured on a percentage scale (Source: Prados de la Escosura (2021)).

	1984-1988	1989-1993	1994-1998	1999-2003	2004-2008
GDP pc	34,227	37,507	41,199	46,552	50,770
Life exp	74.68	75.35	76.14	76.90	77.79
Edu Att	74.74	78.50	81.84	84.06	85.64
Dem Index	0.79	0.81	0.81	0.81	0.81
RLS	5.86	6.07	5.70	6.22	6.57
LS	6.50	6.49	7.10	6.87	7.11
			No shift		
$\hat{m}^*$	0.12	0.12	0.12	0.12	0.12
k^*	0.00	0.00	0.00	0.00	0.00
d	1.79	1.88	2.10	2.35	2.42
h_{min}	0.00	0.00	0.00	0.00	0.00
h_{max}	17.95	18.85	21.01	23.54	24.20
M-LINE$^{k=0}$	11.68	12.23	14.92	16.20	17.21
			No stretch		
$\hat{m}^*$	0.12	0.12	0.12	0.12	0.12
k	3.65	3.96	4.63	5.39	5.57
d^*	1.00	1.00	1.00	1.00	1.00
h_{min}	3.65	3.96	4.63	5.39	5.57
h_{max}	13.65	13.96	14.63	15.39	15.57
M-LINE$^{d=1}$	10.15	10.45	11.72	12.27	12.68

Table A2: Economic and Happiness Indicators for the USA on the long term (cont'd)

Reading note: RLS and LS are the sample mean from the values reported in the survey. $\hat{m}^*$ is a fixed parameter which was calibrated on the US online survey. k and d are, respectively, the shifting and stretching parameters. The asterisk * is used whether the parameter is fixed instead of estimated. h_{min} and h_{max} indicate the location of, respectively, the bottom and the top of the happiness scale, expressed with respect to 1959. M-LINE$^{k=0}$ and M-LINE$^{d=1}$ denote latent satisfaction after the correction for rescaling and under the hypothesis that, respectively, $k = 0$ (no shift) or $d = 1$ (no stretch). Real gross domestic product per capita is measured as chained 2012 Dollars (Source: St.Louis Federal reserve Bank). Life at expectancy at birth is measured in life years (Source: United Nations - World Population Prospects); Educational Attainment at high school is measured in percentage of the population (Source: US Census Bureau). Liberal democracy index is measured on a percentage scale (Source: Prados de la Escosura (2021))

	2017	2018	2019	2020	2022
RLS	7.22	7.29	7.17	6.58	-
LS	7.67	7.59	7.61	7.69	7.44[a]
No shift					
$\hat{m}^*$	0.12	0.12	0.12	0.12	-
k^*	0.00	0.00	0.00	0.00	0.00
d	0.96	0.98	0.96	0.87	1.00
h_{min}	0.00	0.00	0.00	0.00	0.00
h_{max}	9.57	9.78	9.59	8.71	10.00
M-LINE$^{k=0}$	7.33	7.41	7.30	6.71	7.44
No stretch					
$\hat{m}^*$	0.12	0.12	0.12	0.12	
k	-0.33	-0.18	-0.31	-0.99	0.00
d^*	1.00	1.00	1.00	1.00	1.00
h_{min}	-0.33	-0.18	-0.31	-0.99	0.00
h_{max}	9.67	9.82	9.69	9.01	10.00
M-LINE$^{d=1}$	7.33	7.41	7.30	6.71	7.44

Table A3: Happiness Indicators in Germany (2017-2022)

Reading note: RLS and LS are the sample mean from the values reported in the SOEP-IS survey. $\hat{m}^*$ is a fixed parameter which was calibrated on the US online survey. k and d are, respectively, the shifting and stretching parameters. The asterisk * is used whether the parameter is fixed instead of estimated. h_{min} and h_{max} indicate the location of, respectively, the bottom and the top of the happiness scale, expressed with respect to 2022. M-LINE$^{k=0}$ and M-LINE$^{d=1}$ denote latent satisfaction after the correction for rescaling and under the hypothesis that, respectively, $k = 0$ (no shift) or $d = 1$ (no stretch). [a] LS_{2022} is not representative of the national population, since only one-fourth of the SOEP-IS sample answered the standard question on life satisfaction that year. All other mean estimates are representative.

	2016	2017	2018	2019	2020	2021	2022	2023
RLS	6.33	6.38	6.35	6.39	6.72	5.96	6.24	6.34
LS	6.52	6.61	6.48	6.60	6.72	6.55	6.52	6.52
				No shift				
$\hat{m}^*$	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12
k^*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
d	1.00	1.00	1.02	1.02	0.98	1.09	1.12	1.13
h_{min}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
h_{max}	10.00	10.03	10.23	10.19	9.83	10.88	11.21	11.30
M-LINE$^{k=0}$	6.52	6.63	6.63	6.72	6.61	7.13	7.30	7.36
				No stretch				
$\hat{m}^*$	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12
k	0.00	0.02	0.15	0.13	-0.12	0.53	0.72	0.78
d^*	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
h_{min}	0.00	0.02	0.15	0.13	-0.12	0.53	0.72	0.78
h_{max}	10.00	10.02	10.15	10.13	9.88	10.53	10.72	10.78
M-LINE$^{d=1}$	6.52	6.63	6.63	6.72	6.61	7.09	7.24	7.29

Table A4: Happiness Indicators in France (2016-2023)

Reading note: RLS and LS are the sample mean from the values reported in the CAMME survey. $\hat{m}^*$ is a fixed parameter which was calibrated on the US online survey. k and d are, respectively, the shifting and stretching parameters. The asterisk * is used whether the parameter is fixed instead of estimated. h_{min} and h_{max} indicate the location of, respectively, the bottom and the top of the happiness scale, expressed with respect to 2016. M-LINE$^{k=0}$ and M-LINE$^{d=1}$ denote latent satisfaction after the correction for rescaling and under the hypothesis that, respectively, $k = 0$ (no shift) or $d = 1$ (no stretch).

	2018	2023
RLS		6.56
LS	4.66	4.67
No shift		
$\hat{m}^*$		0.12
k^*	0	0
d	1	0.69
h_{min}	0	0
h_{max}	10	6.98
M-LINE$^{k\,=\,0}$	4.66	3.26
No stretch		
$\hat{m}^*$		0.12
k^*	0	-2.01
d	1	1
h_{min}	0	-2.01
h_{max}	10	7.98
M-LINE$^{d\,=\,1}$	4.66	2.66

Table A5: Happiness Indicators in Ukraine (2018-2023)

Reading note: RLS and LS are the sample mean from the values reported in the Gallup World Poll survey. $\hat{m}^*$ is a fixed parameter which was calibrated on the US online survey. k and d are, respectively, the shifting and stretching parameters. The asterisk * is used whether the parameter is fixed instead of estimated. h_{min} and h_{max} indicate the location of, respectively, the bottom and the top of the happiness scale, expressed with respect to 2016. M-LINE$^{k\,=\,0}$ and M-LINE$^{d\,=\,1}$ denote latent satisfaction after the correction for rescaling and under the hypothesis that, respectively, $k = 0$ (no shift) or $d = 1$ (no stretch).

A3 Survey procedures

A3.1 US online panel

The survey was programmed using Qualtrics and participants were recruited on Prolific (Palan and Schitter, 2018). Data collection took place on January 3, 2024 (Wave 1) and January 17-20, 2024 (Wave 2). Subjects were recruited to form a representative sample of the U.S. population based on Census composition by gender and age. Age and gender composition of the sample are reported in table A6. They were informed that they would be recontacted for a second survey at the very beginning of Wave 1. Median completion time was around 2 minutes and participation paid $0.4 (i.e., 12 $/hour rate). All of the participants gave their informed consent at the beginning of the study. 1,997 participants responded to the first wave of the survey, while 1,648 responded to the second wave too (17% attrition). We excluded 20 participants who spent, on average, less than a minute on the survey, i.e., the first percentile of the distribution of the response time. The analytical sample is a balanced panel of 1,628 respondents.

	N	mean	sd	min	max
age	1,628	47.30	15.50	20	80
male	1,628	.49	.50	0	1

Table A6: Age and gender composition of the US online sample

A3.2 UK online panel

The survey was programmed using Qualtrics and participants were recruited on Prolific (Palan and Schitter, 2018). Data collection took place in February 2023 (Wave 1) and February 2024 (Wave 2). Subjects were all UK resident and were informed that they would be recontacted for a second survey at the very beginning of Wave 1. Age and gender composition of the sample are reported in table A7. Participants received £1.5 and £1 for their participation to, respectively Wave 1 (7 minutes) and Wave 2 (5 minutes). These payments correspond to hourly rates above £12. All of the participants gave their informed consent at the beginning of the study. 1,502 participants responded to the first wave of the survey, while 1,023 responded to the second wave too (32% attrition). We excluded 13 participants who failed an attention check and 4 that we failed to match based on our records. Further, 19 participants were considered as outliers based on their level of rescaling, according to the criteria described in section 5.1. The analytical sample is a balanced panel of 987 respondents. Additional information about the data collection is available in Kaiser and Prati (2024).

	N	mean	sd	min	max
age	987	42.95	13.27	18	82
male	987	0.49	.50	0	1

Table A7: Age and gender composition of the UK online sample

A4 Model details

The core of the model rests in two equations. The first one is the modeling of life satisfaction (LS_t) as a linear mapping of latent satisfaction (h_t) into a 0-10 space regulated by a shifting parameter (k_t) and a stretching parameter (d_t), defined with respect to a baseline year.

$$LS_t = \frac{h_t - k_t}{d_t} \tag{A1}$$

The second one is the modeling of retrospective life satisfaction ($RLS_{t,t-1}$) as a linear mapping of past latent satisfaction (h_{t-1}) on the current life satisfaction scale, as defined by the scaling parameters (k_t and d_t).

$$RLS_{t,t-1} = \frac{h_{t-1} - k_t}{d_t} - \bar{m} \tag{A2}$$

Combining equations (A1) and (A2), we obtain the general recursive form of $h_t \equiv$ M-LINE$_t$:

$$\begin{aligned} LS_t - RLS_{t,t-1} &= \frac{h_t - k_t - h_{t-1} + k_t}{d_t} + \bar{m} \\ d_t(LS_t - RLS_{t,t-1}) &= h_t - h_{t-1} + d_t\bar{m} \\ h_t &= h_{t-1} + d_t(LS_t - RLS_{t,t-1} - \bar{m}) \end{aligned} \tag{A3}$$

A4.1 No-stretch and no-shift scenarios

In the paper, we consider two restrictions.

If $d = 1$ (no stretch), then:

$$LS_t = h_t^{d=1} - k_t \tag{A4}$$

$$RLS_{t,t-1} = h_{t-1}^{d=1} - k_t - \bar{m} \tag{A5}$$

Solving for k_t, yields the value of $h_t^{d=1} \equiv$ M-LINE$_t^{d=1}$:

$$\begin{aligned} LS_t - h_t^{d=1} &= RLS_{t,t-1} - h_{t-1} + \bar{m} \\ h_t^{d=1} &= h_{t-1}^{d=1} + (LS_t - RLS_{t,t-1} - \bar{m}) \end{aligned} \tag{A6}$$

If $k = 0$ (no shift), then:

$$LS_t = \frac{h_t^{k=0}}{d_t} \tag{A7}$$

$$RLS_{t,t-1} = \frac{h_{t-1}^{k=0}}{d_t} - \bar{m} \tag{A8}$$

Solving for d_t, yields the value of $h_t^{k=0} \equiv \text{M-LINE}_t^{k=0}$:

$$\frac{h_t^{k=0}}{LS_t} = \frac{h_t^{k=0}}{RLS_{t,t-1} + \bar{m}}$$
$$h_t^{k=0} = h_{t-1}^{k=0} \cdot \frac{LS_t}{RLS_{t,t-1} + \bar{m}} \tag{A9}$$

A4.2 Conditions under which the no-stretch and no-shift scenarios are boundaries of latent satisfaction

Latent satisfaction is comprised between $\text{M-LINE}_t^{d=1}$ and $\text{M-LINE}_t^{k=0}$ insofar as:
Condition 1:

$$h_t^{d=1} \leq h_t$$
$$h_{t-1} + (LS_t - RLS_{t,t-1} - \bar{m}) \leq h_{t-1} + d_t(LS_t - RLS_{t,t-1} - \bar{m}) \tag{A10}$$

That is verified when:

$$\text{Either} \quad LS_t \geq RLS_{t,t-1} + \bar{m} \quad \& \quad d_t \geq 1; \tag{A11}$$
$$\text{or} \quad LS_t \leq RLS_{t,t-1} + \bar{m} \quad \& \quad d_t \leq 1 \tag{A12}$$

Condition 2:

$$h_t^{k=0} \geq h_t$$
$$h_{t-1} \cdot \frac{LS_t}{RLS_{t,t-1} + \bar{m}} \geq h_{t-1} + d_t(LS_t - RLS_{t,t-1} - \bar{m}) \tag{A13}$$

That is verified when:

$$\text{Either} \quad LS_t \geq RLS_{t,t-1} + \bar{m} \quad \& \quad d_t \leq \frac{h_{t-1}}{RLS_{t,t-1} + \bar{m}}; \tag{A14}$$
$$\text{or} \quad LS_t \leq RLS_{t,t-1} + \bar{m} \quad \& \quad d_t \geq \frac{h_{t-1}}{RLS_{t,t-1} + \bar{m}} \tag{A15}$$

By replacing $RLS_{t,t-1} + \bar{m}$ with its expression from eq.A2, we obtain:

$$\frac{h_{t-1}}{RLS_{t,t-1} + \bar{m}} = d_t$$
$$\frac{h_{t-1}}{\frac{h_{t-1}-k_t}{d_t} - \bar{m} + \bar{m}} = d_t$$
$$d_t \cdot \frac{h_{t-1}}{h_{t-1} - k_t} = d_t \tag{A16}$$

Which is verified iff $k_t = 0$. Therefore, *Condition 2* simplifies as follows:

$$\text{Either} \quad LS_t \geq RLS_{t,t-1} + \bar{m} \quad \& \quad k_t \geq 0; \tag{A17}$$
$$\text{or} \quad LS_t \leq RLS_{t,t-1} + \bar{m} \quad \& \quad k_t \leq 0 \tag{A18}$$

By combining *Conditions 1* and *2*, we obtain:

$$\text{Either} \quad LS_t \geq RLS_{t,t-1} + \bar{m} \quad \& \quad d_t \geq 1; \ \& \quad k_t \geq 0; \tag{A19}$$
$$\text{or} \quad LS_t \leq RLS_{t,t-1} + \bar{m} \quad \& \quad d_t \leq 1; \& \quad k_t \leq 0 \tag{A20}$$

By using the fact that $k_t \equiv h_{min}$ and $d_t \equiv \frac{h_{max}-h_{min}}{10}$, the conditions can also be rewritten as:

$$\text{Either} \quad LS_t \geq RLS_{t,t-1} + \bar{m} \quad \& \quad 0 \leq h_{min} \leq h_{max} - 10; \tag{A21}$$
$$\text{or} \quad LS_t \leq RLS_{t,t-1} + \bar{m} \quad \& \quad 0 \geq h_{min} \geq h_{max} - 10 \tag{A22}$$

That is, when people feel better than yesterday, the scale can be stretching and/or shifting upwards. When people are feeling worse, the scale can be shrinking and/or shifting downwards. Said differently, when people feel better than yesterday, their idea of best possible life must improve at least as fast as their idea of worst possible life does; when people feel worse than yesterday, their idea of worst possible life must deteriorate at least as fast as their idea of best possible life. Overall, M-LINE$_t^{k=0}$ and M-LINE$_t^{d=1}$ represent the lower and upper-bounds of latent satisfaction insofar as the scale stretches upwards or shrinks downwards. The table below summarizes the conditions:

	Shift Up ($k_t > 0$)	**No Shift** ($k_t = 0$)	**Shift Down** ($k_t < 0$)
Stretch ($d_t > 1$)	✓	✓	x
No Stretch ($d_t = 1$)	✓	✓	✓
Shrink ($d_t < 1$)	x	✓	✓

Table A8: Conditions under which M-LINE$_t^{k=0} \leq h_t \leq$ M-LINE$_t^{d=1}$.