NET WORK LISER

Discussion Paper Series

IZA DP No. 18536 April 2026

Is It Possible to Raise National Happiness?

Alberto Prati
University College London, University of Oxford and LSE

Claudia Senik
Sorbonne University, Paris School of Economics, IUF and IZA@LISER

IZA DP No. 18536 April 2026

Is It Possible to Raise National Happiness?

Abstract

We revisit the famous Easterlin paradox by considering that life evaluation scales refer to a changing context, hence they are regularly reinterpreted. We propose a simple model of rescaling based on both retrospective and current life evaluations, and apply it to unexploited archival data from the USA. When correcting for rescaling, we 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. Using several datasets, we shed light on other happiness puzzles, including the apparent stability of life evaluations during COVID-19, why Ukrainians report similar levels of life satisfaction today as before the war, and the absence of parental happiness.

JEL classification

I31, N32, O10

Keywords

happiness, life satisfaction, subjective well-being, rescaling, Easterlin paradox, Cantril ladder, Gallup

Corresponding author

Claudia Senik
claudia.senik@psemail.eu

"Raising the incomes of all does not increase the happiness of all. This is because the material norms on which judgments 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 remain 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 well-being framework, 85% use life satisfaction (Mahoney, 2023). In the U.K., 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 a 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 (henceforth LS) is flat.[1] 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.[2]

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

[1] This stylized fact constitutes the ink bottle of the economics of happiness. Although disputed by some scholars (Stevenson and Wolfers, 2008; Sacks et al., 2012), it has been shown for a large set of countries (Easterlin, 1974, 1995, 2001; Easterlin and O'Connor, 2020; Easterlin et al., 2010). It is illustrated by the time series published by the World Happiness Report on the basis of the Gallup World Poll, in the case of most developed countries (Western Europe, North America, and New-Zealand) and several rapidly growing economies (like India or Bangladesh), but not in Central and Eastern European countries or China (Helliwell et al., 2025).

[2] 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 becoming 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.

that they were worse off in the past and will be better off in the future, although their reports on present happiness remain 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 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 to, e.g., by the initial Cantril formulation, as "the worst possible life" (0) and "the best possible life for you" (10).[3] 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 as follows: What is context-dependent: latent experience or the measurement scale? What is the underlying mechanism that keeps the trend in reported happiness flat: *hedonic treadmill* or *rescaling*?[4]

Proponents of the *hedonic treadmill* interpretation, following Richard Easterlin, argue that life satisfaction is inherently relative, shaped by adaptation (Easterlin, 1974, 1995, 2001; Di Tella and MacCulloch, 2010). The adaptation hypothesis extends neurobiological homeostasis—as established by neurophysiological research—to the domain of life satisfaction.[5] This framework aligns with Kahneman and Tversky's prospect theory (1979), which posits that experienced happiness depends not on absolute levels (e.g., of income), but on deviations from reference points. These reference points evolve through temporal adaptation, whereby past experiences recalibrate aspirations. The fact that intertemporal judgments do not match concurrent judgments, in turn, can be explained as a memory illusion, or a "focusing illusion" (Kahneman et al., 2006): when asked to reconstruct their past satisfaction, respondents assess their past circumstances instead of their past feelings (Easterlin, 2001).

By contrast, partisans of *rescaling* have pointed out that the happiness scale is naturally context-

[3]The formulation of the Cantril ladder question is the following: "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?".

[4]Other explanations of the paradox point to income comparisons (Clark and Senik, 2010; Clark, 2016; Clark et al., 2008), limits to the hedonic return on material affluence (Layard, 2005; Kahneman and Deaton, 2010), growing inequality of income distribution (Alesina et al., 2004; Oishi et al., 2011), and the large hedonic cost of short recessions (De Neve et al., 2018).

[5]Direct observations of the brain's evaluative systems reveal two key adaptation mechanisms: (1) homeostatic regulation that maintains emotional equilibrium (Kleckner et al., 2017; Barrett, 2017), and (2) measurable neural adjustments in reward circuitry, particularly downregulation of dopamine receptors to prevent overstimulation (Volkow et al., 1993) and context-dependent coding of rewards (Tobler et al., 2005).

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 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 it correctly, but they measure it on their current scale, which differs from the one used at the epoch.

Formally, the fundamental difference between the two conjectures is that the hedonic treadmill assumes a time-dependent *utility* function, while rescaling assumes a time-dependent *reporting* function.[6] 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 definitively separate 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 attempt to quantify the role of each of them.

Rather, we focus on the second one and show what a reconstructed time series of latent life satisfaction would be like under the rescaling hypothesis, based on a new method that exploits current and retrospective satisfaction. We call this reconstructed series the M-LINE. Most of the existing literature has interpreted life satisfaction time series from the perspective of the 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 and build the M-LINE. 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 clear instructions to survey data stakeholders about how to use this method for future longitudinal data.

Second, we 'reconstruct' the evolution 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 conducted by Cantril in 1959 and ending in 2008. We show that, depending on the assumption about the type of rescaling process, we obtain a life satisfaction index (represented by the M-LINE) that evolves more or less closely in parallel with GDP per capita and the Human Development Index. The Easterlin paradox, therefore, disappears.

[6] In the paper, we use the term "utility" in the sense of ex-post "experienced utility" à la Kahneman - intended as subjective hedonic experience - and not ex-ante "decision utility" - intended as the weight of an outcome in a decision (Kahneman et al., 1997). This use is consistent, for instance, with the excerpt from Kahneman and Krueger (2006)'s discussion of rescaling that we report at p.9 ("if people gradually adjust their aspirations to the utility that they normally experience..."), but should be distinguished from the use of "utility" in other papers on subjective well-being. For instance, Benjamin et al. (2012) show that maximizing happiness only explains choices in 80% of cases, and conclude that "people do not seek to maximize SWB exclusively, at least as it is currently measured, but that SWB is a uniquely important argument of the utility function" (p.2017).

Third, when we apply our 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.[7] 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 nominal life satisfaction remained unchanged in 2023 compared to 2018. However, our index reveals that the Russian invasion had a significant negative impact on the latent life satisfaction of Ukrainians - an effect that remains hidden when rescaling is overlooked.

We also 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 the 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 the same whether using LS or our new index as dependent variables, except for the coefficient associated with childbirth, which flips to a positive value when accounting for rescaling. Finally, we find that most respondents perceive some evolution in what they mean by "0" and "10" on the LS scale, suggesting that rescaling is indeed a genuine process.

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; Kaiser and Lepinteur, 2025). Arguably, the papers closest to ours are Kaiser (2022) and Benjamin et al. (2023b). Kaiser (2022) was the first to propose using retrospective satisfaction reports to correct for the 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) model the relationship between latent LS and reported LS using a similar conceptual framework as we do, but they are interested in scale-use heterogeneity *across individuals*, rather than *over time*. They propose some *calibration questions* to adjust for potential differences in the respondents' interpretation of the LS scale.

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 - the M-LINE - and provides some instructions to survey data stakeholders. Section 6 concludes.

[7] Life satisfaction from national panel data displays a flat trend too. This can be observed in the U.K. Understanding Society Survey panel, in the Household Income and Labor Dynamics panel in Australia, and in the German Socio-Economic Panel.

1 Literature

The question of 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, which 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 welfare. 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., 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 respondents 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 latent 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, has been discussed and tested.[8] In a recent paper, Kaiser (2022) reported a number of previous experiments showing

[8]Some authors (e.g. Fabian, 2022b) use the term "scale norming" to refer to what we call here "rescaling". Other

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).

However, it is difficult to distinguish a shift in the scale from a shift in actual satisfaction. This applies to the "preference drift" uncovered by the Leyden school. This expression refers to the change in people's aspirations as their own income increases (Van Praag, 1971), where income aspirations are measured by an individual's answer to the Minimum Income Question.[9] 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 and higher reported position compared to their pre-immigration status, successful migrants did not report higher well-being scores, a contradiction that the authors attribute to rescaling.

However, again, these results can be interpreted either as reflecting the hedonic treadmill or as 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).[10] For instance, people who were victims of spinal cord injury have been shown to use a quality-of-life scale quite differently in the five years following the accident (Schwartz et al., 2018).[11] The seminal study of this strand of literature is Brickman et al. (1978), which showed that people who became paraplegic after an accident rated 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

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).

[9] "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."

[10] The widely cited idea that victims of severe accidents quickly and completely recover their happiness scores has 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).

[11] Schwartz et al. (2018) follows the approach developed by Oort et al. (2005) and measures a change in the use of the scale by studying the change in conditional correlations between life satisfaction and observable 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).

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 designed to address the two issues are closely related. 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 in the population regarding how they use subjective well-being scales. Our approaches are complementary, as we do not address, in the present study, the issue of interpersonal comparability of life satisfaction reports.

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.

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 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 by the same magnitude, latent satisfaction h remains unchanged.[12]

Easterlin also mentions another possible reason for the flatness of the self-declared nominal level of satisfaction - 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

[12]"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).

average happiness remains unchanged (zero-sum game at best, negative-sum game if comparisons are only upward).[13]

Concerning the second stylized fact, that people think they were less happy in the past and will be happier in the future, Easterlin's interpretation is again based on the hedonic treadmill hypothesis: "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" (Easterlin, 2001, p.465).

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 aspirations. In that framework, individuals employ memories of objective drivers of happiness c_{t-1} rather than of latent happiness h_{t-1} itself.[14]

Note that Easterlin does not make a distinction between the latent life satisfaction felt by people (h) and their nominal level of satisfaction reported 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 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 relation between latent happiness and 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 between the two views can be formalized as follows:

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 LS 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.:

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

[13] "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).

[14] 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).

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 $\bar{c}$. Hence, the well-being effect of a change in personal circumstances is separable from a change in the reference point.

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 onto the proposed 0-10 scale, which is described as the scope of possibilities spanning from "the worst possible life" (h^{min}) to "the best possible life for you" (h^{max}). As this scope changes over time, so does the signification of a given step on the scale. Hence, latent satisfaction h can be rising although reported 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 intertemporal judgments of life satisfaction, and explains that "we may need to adjust concurrent measures of happiness for inflation of standards over time".[15]

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, for instance). In order to calculate the desired index of latent satisfaction h over time, one must perform 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.

An important point is that, to follow this route, 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.

Our objective is to define the trajectory of latent happiness h_t relative to a fixed point in time (a numeraire). Without loss of generality, we define this reference point as the reported level of life satisfaction (LS) at the first period of study ($t = 0$), i.e.:

$$h_0 \equiv LS_0 \tag{3}$$

At t_0, the "worst possible life" is, by construction, indicated by the lowest rung (0) of the LS scale, and the "best possible life" by the highest rung (10). Consistent with eq.(3), we use the latent scale at t_0 as our numeraire, so that $h_0^{\min} = 0$ and $h_0^{\max} = 10$.

Then, as latent happiness changes over time, and subjects still have to report it on the same 0-10

[15] "[...] 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" (Hagerty, 2003, p.136).

LS scale, we assume that they might reinterpret the minimum rung 0 as corresponding to the new level of happiness in the "worst possible life" and the maximum rung 10 as the happiness level in the "best possible life". In other words, the scope of latent happiness levels (from the worst possible life to the best possible life) might change, while respondents still have to project it on the same 0-10 step LS scale [16]. We consider two possible ways in which the latent happiness scale (i.e., the set of possibilities) may have evolved: (i) upward or downward shifting; (ii) stretching or shrinking.

In the case of a shift, the position of the scale has translated upward (respectively downward), meaning that the floor of the scale associated, i.e. the "worst possible life", has increased (resp. decreased) by k_t:

$$k_t = h_t^{min} \tag{4}$$

In the case of a stretch (respectively shrink), the initial latent happiness scale expands (resp. contracts) by a factor d_t in time t as compared to time t_0, meaning that the "best possible life" is now d_t times higher (resp. lower) compared to the worst possible life. As subjects still have to report h_t on the 0-10 LS scale, we have:

$$d_t = \frac{h_t^{max} - h_t^{min}}{10} \tag{5}$$

Hence, the reporting function can be modeled as an affine transformation of the latent happiness, governed by a shifting parameter k_t and a stretching parameter d_t, i.e.:

$$LS_t = h_t \times \frac{1}{d_t} - \frac{k_t}{d_t} \tag{6}$$

For example, if the latent scale stretches because d changes from 1 to 1.1 (and k=0 does not change), a latent life satisfaction of 8 will become a reported score of 7.3. If the latent scale shifts upward by k=2 (and d=1 does not change), then a latent life satisfaction of 8 will become a reported score of 6.

This equation is key. It describes the mapping of the latent feeling h_t into the 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.

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

[16] This may be due to social comparisons, as people learn about "the best possible life" or "the worst possible life" from observing other individuals or groups in society. However, we need not assume that these observations exert a direct hedonic effect on their satisfaction (through jealousy or frustration), it is sufficient to assume that it modifies their reporting function.

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:

$$RLS_{t,t-1} = \frac{h_{t-1}}{d_t} - \frac{k_t}{d_t} \tag{7}$$

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 being due to rescaling and use this information to convert h_t and h_{t-1} into comparable units:

$$h_t = h_{t-1} + (LS_t - RLS_{t,t-1}) \times d_t \tag{8}$$

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

$$h_t = LS_0 + \sum_{n=1}^{t} (LS_n - RLS_{n,n-1}) \times d_n \tag{9}$$

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 identification restriction. As we will see, this 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.[17]

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: $k_t = k_0 = 0$, $\forall t$. The scope of possibilities can then grow ($d > 1$) or shrink ($d < 1$), and the

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

series of reconstructed h_t will write: $h_t = d_t \times LS_t$. Therefore:

$$h_t = LS_t \times \prod_{n=1}^{t} \frac{LS_{n-1}}{RLS_{n,n-1}} \tag{10}$$

Where, at a given point in time t:

$$LS_{t-1}/RLS_{t,t-1} = d_t/d_{t-1} \tag{11}$$

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 $d_t = d = 1,\ \forall t$. Authors such as Fabian (2022b) focus on this scenario, where the difference between the best and the worst possible life remains constant. The set of possible levels of h can however 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:

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

And at a given point in time t:

$$LS_{t-1} - RLS_{t,t-1} = k_t - k_{t-1} \tag{13}$$

2.6 Raising the happiness of all or raising aggregate happiness

When we use aggregate cross-sections (section 4), we assume the meaning of the scale is determined by the best/worst possible outcomes for everyone. In that case, comparisons *within* a country are not considered.[18] The long time-series that we use is not a panel, as will be described in the data section, so we are restricted to studying the evolution of aggregate latent satisfaction, here average satisfaction. Our index thus refers to a "representative" or aggregate agent at the country level. Notice that in both cases of no shifts or no stretch, if the parameters d_t and k_t are the same across individuals, the equations that describe the dynamics of latent happiness h_t are linear in the number of individuals, so they apply to individual happiness as well as to the average happiness of a country's inhabitants. [19] If we had *long* panel data containing information about current and retrospective LS for each respondent, we could compute an index of reconstructed latent happiness for each individual and take into account individual circumstances. To the best of our knowledge, such a long panel

[18] Instead, comparisons *between* countries are not assumed away, since aggregating each national cross-section into one observation corresponds to a representative-agent framework, where each country's population is treated as an agent.

[19] In theory, we could recover information about individual heterogeneity also from repeated cross-sections, but only at the cost of heroic assumptions about the variance of the memory error - section A5, in the Appendix, presents an extension that illustrates this approach.

does not exist. However, *short* panels (up to four years) do.

We use such short-run individual-level panel data in section 5. In this case, we can calculate the rescaling parameters k_{it} and d_{it} for each individual i at time t. Section 5.4 will discuss some results about rescaling heterogeneity. It is fully possible that these individual-level parameters are affected by inter-personal comparisons, i.e. that different people may have different interpretations of the LS scale, depending on their own idea of the worst and best possible lives, maybe influenced by their reference group. We are not able to measure these relative concerns, but we do not need to assume them away.

2.7 Memory correction

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, and 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.

In this paper, we assume that the memory bias tends to be constant. Indeed, although the literature on memory retention shows that recall accuracy decreases over time (as predicted by the law of recency), it 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. $m_{t,t-n} \sim D(\bar{m}, v_n)$.[20] This assumption is crucial for our framework. It implies that when respondents recall their past satisfaction, they will, on average, be "off" by a certain amount, regardless of how long ago the recalled state occurred - be it one, two, or five years ago.

Although this assumption is grounded in the cognitive psychology literature, we are not aware of a previous experiment that tested this scenario under ideal conditions. In an ideal experiment, respondents should report a latent *continuous* variable on a short *ordered* scale, and, later on, give a retrospective assessment of the latent for different retrospective intervals. The scale must be *objective*, to isolate the recall bias, without the possibility of rescaling as a confounder.

We run such an experiment, using income as the latent variable and income brackets as the ordinal objective scale. In March 2025, we asked respondents of a UK online panel (described in Section 3.3) to report what their monthly gross income was twelve months ago and two years ago, by choosing their corresponding income bracket. We compare their retrospective answers to the answers they gave at the epoch - in 2023 and 2024 - and estimate the average recall error for a 12-month and 24-month retention interval. We code the income brackets from 0 (below £500) to 16 (above £8,000). We find that, whether the retrospective question refers to one or two years ago, the

[20] The variance of the memory error does not play a role in our representative agent model, but it can play a pivotal role in a model that exploits individual heterogeneity (see Appendix, section A5).

average recall error is statistically the same (N=557, diff=-0.03, p-value = 0.636) - see the Appendix, fig.A2. If we top-code the higher income brackets and reduce the number of categories to eleven so that we match the range of the satisfaction scale, results are unchanged (N=557, diff=-0.01, p-value = 0.912) - see the Appendix, fig.A3. Overall, this empirical exercise supports the assumption of a constant memory bias.

2.8 Memory bias estimation

Based on the assumption that the memory bias is constant (as discussed in the previous section), we now try to estimate its value in the case of life satisfaction. To this end, we need to isolate it from rescaling. The key of our strategy is to look at retrospective satisfaction over a very short time interval over which distortions of the reporting functions are likely to be negligible. By contrast, 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 interviewed 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 a 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 life satisfaction is, on average, *lower* than current life 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: [0.06;0.17]). This estimate is very similar to the one obtained by Anvari (2023), who ran a similar exercise with respondents from various countries. In his 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: [0.08;0.19], N=1,332).[21]

2.9 The M-LINE

Based on our external calibration exercise, we 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).[22] Note that the results are very similar if we do not correct for this memory error parameter (see the Appendix, fig.A4). Section A4.3, in the Appendix, reports the details of the derivation of equations (14) and (15).

[21] 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 (2023) finds 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).

[22] 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).

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

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

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:

$$h_t = LS_t + \sum_{n=1}^{t} (LS_{n-1} - RLS_{n,n-1}) + t\bar{m}. \tag{15}$$

We will refer to latent satisfaction calculated under the assumption of no shift (eq. (14)) as M-LINE$^{k=0}$. For latent satisfaction with no stretch (eq. (15)) 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. This measure is uniquely collected around the globe every year; it is analyzed and discussed in the annual World Happiness Report. It has been suggested that, compared to other measures of life satisfaction, the Cantril ladder tends to elicit thoughts related to affluence (Nilsson et al., 2024), which makes the Easterlin paradox even more salient.

Our exercise requires historical records on both current and retrospective life satisfaction, measured on a cardinal scale. Both measures were originally introduced by Cantril (1965). (We interchangeably use the terms life evaluation and life satisfaction to refer to the self-reported step indicated by respondents on the scale). The first poll was run by Cantril's team in 1959 and continued by Gallup inc. until today; however, to date, the full evolution of the Cantril ladder has never been reconstituted. To do so, we used recently digitized archival data from the Roper Center at Cornell University and merged it with unexplored data from Gallup US polls in the early 2000s and the Gallup World Poll. This is the first reconstitution of long time series (1959-2008) containing both current and retrospective life evaluations.

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 USA 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

Reading note: Survey-weighted estimated means, based on representative samples of the USA population. In 1959, the average current LS was 6.58, while the mean retrospective LS (referring to five years earlier) was 5.89. Five years later, in 1964, the LS of 1959 was recalled as 5.92. RLS wording: "On which step would you say you stood five years ago?" [0-10]. *Only sample means 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, but 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 in time, in the same survey wave. The difference is systematically positive and is, in fact, 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).[23] 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 analyzes 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 A4-A10.

3.2 Short-run times series (cross-section)

We also use two series of repeated 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 4.2.2.

The second source is the aforementioned Gallup World Poll. We leverage a new question on the retrospective Cantril ladder that was added to the questionnaire in 2023 and that was fielded in Ukraine. This data set will be analyzed in section 4.2.3.

3.3 Short-run times series (panel)

Panel data provide the ideal setting 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 as of 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).[24] Before our study, this was the only extant large panel dataset that measured RLS. We will analyze this data in section 5.1 and 5.3.

[23] Concretely, we consider each 5-year stretch as one period, and average the data within each period. For instance, we take the average retrospective satisfaction reported in 2004, 2005, 2006, 2007 and 2008 (i.e., period 2004-2008) and compare it to the satisfaction reported in 2001, 2002 and 2003 (i.e., period 1999-2003).

[24] Odermatt and Stutzer (2019) show that people's predictions of their future happiness following a major life event are systematically inaccurate, stemming from adaptation. In a robustness test, they look at the gap between LS_{t-1} and $RLS_{t,t-1}$, using a sub-sample of data and conclude that rescaling cannot fully explain the life satisfaction trajectory. Klemm (2022) detects a systematic mismatch of the time series of current reported LS with respect to retrospective and prospective assessments of LS, and highlights the usefulness of using intertemporal assessments of SWB.

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 trajectory of nominal life satisfaction reported in each year t (i.e., LS_t). We will analyze it in section 4.2.1.

The third panel is an online sample, representative of the American population, that we interviewed twice, at two-week distance, in January 2024. We briefly describe this panel in section 2.8. It is a balanced panel of 1,627 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 "U.S. 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-March 2023 responded again one year later, in February-March 2024.[25] 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 short-run datasets that we use.

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, estimated on the basis of LS_t, LS_{t-1} and $RLS_{t,t-1}$. The blue line displays the evolution of nominal satisfaction LS_t. The green dashed line indicates the level of satisfaction for the same corresponding year that will be reported retrospectively in 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

[25] 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	Design	Repres.	t	N interviews
CAMME	France	2016-2023	cross-section	✓	1	53,918
Gallup World Poll	Ukraine	2018, 2023	cross-section	✓	5	1,954
SOEP core	West Germany	1984-1987	panel	✓	1	27,485
SOEP-IS	Germany	2017-2022	panel	✓	1-5	6,742
U.S. online panel	USA	2024	panel	✓	5	3,254
U.K. online panel	United Kingdom	2023-2024	panel		1	1,980

Reading note: Overview of six datasets used for the short-term analyses. Column t details the time frame of RLS (in years). RLS wordings are detailed here. CAMME: "And when you think about last year, how did you place on a scale of 0 to 10?" [0-10]. Gallup World Poll: "On which step would you say you stood five years ago?" [0-10]. SOEP core: "How satisfied were you a year ago with your life?" [0-10]. SOEP-IS: "All things considered, how satisfied with your life were you in each of the following years? 2017, 2018, 2019, 2020" [0-10]. U.K. online panel: "And how satisfied were you with your life about one year ago, all things considered?" [0-100]. U.S. online panel: "On which step would you say you stood two weeks ago?" [0-10] "And on which step would you say you stood five years ago?" [0-10].

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 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 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 USA 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. 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. The notion of the "best possible life" (represented by the top-end of the scale) has substantially improved, 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, similar to GDP per capita. The end-point, in 2008, is about 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 response would be 17.2. This represents a markedly different dynamic from what nominal satisfaction would suggest. The interpretation of the top of the scale changed dramatically as well. We estimate that the latent satisfaction associated with reporting the highest step (or any step) of the LS ladder was almost two and a 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 it did 50 years earlier.

The bottom pane of figure 1 displays the evolution of latent well-being in the case where the scale can shift but not stretch/shrink over time ($d = 1$). In this case too, the M-LINE displays a substantial increase in latent satisfaction, which roughly doubled between 1959 and 2008. The assumption that the scale simply shifted upward (without any stretch) may not be very plausible, 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 a scale shift of more than 5 points ($k_{1959} = 0; k_{2008} = 5.57$IQ3).

In the Appendix, tables A1-A2 report the data underpinning figure 1, while figures A4-A10 repeat the analyzes with a memory bias parameter ranging 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.

Note that the period covered by the data ends in 2008, hence at the onset of the subprime crisis, and that it may not capture the recent deterioration in the mental health of Americans, as documented by Case and Deaton (2020).

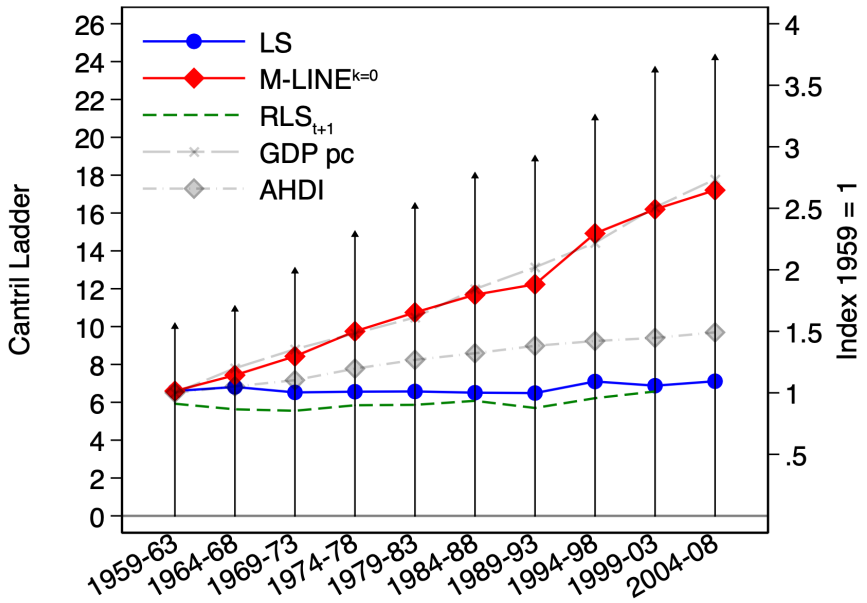
4.2 The M-LINE during dire times: the case of the pandemic in Europe and of the war in Ukraine

The previous section showed that when rescaling is taken into account, the evolution of latent life satisfaction in the USA follows 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 life satisfaction (LS) or can it be below?

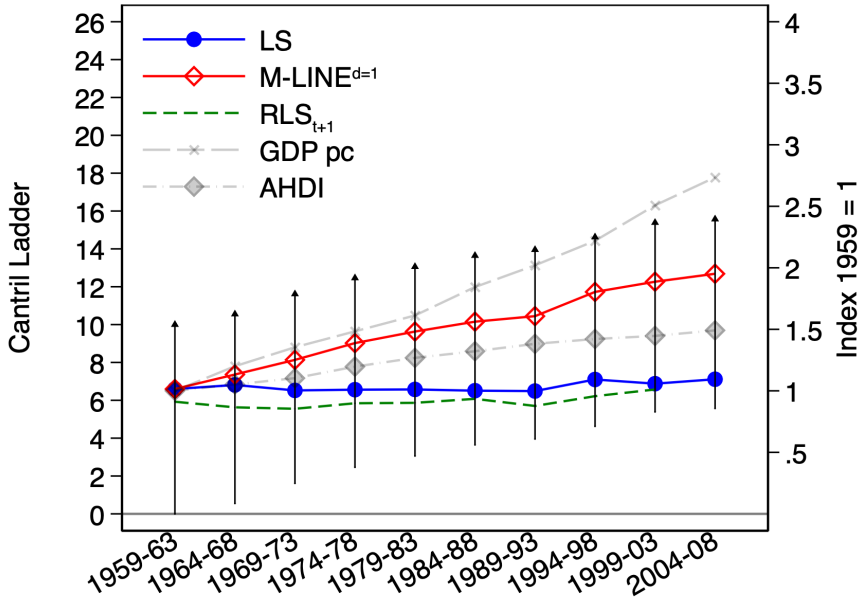
To find out, we study the evolution of the latent scale during dire times, using three datasets. The first one allows reconstructing *a posteriori* the dynamics of latent satisfaction during the COVID-19 pandemic in Germany. The second one leverages retrospective life satisfaction questions asked during 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 graphically presented in this section.

4.2.1 The pandemic in Germany

In Germany, nominal life evaluation 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



(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. $\bar{m} = 0.12$. 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).

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-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.[26]

The level of life satisfaction reported in 2020, the first year of the pandemic, is intriguing, as it is *higher* than in previous periods. However, the retrospective evaluations of that period indicate that it was significantly *worse* than in previous years. In the framework of rescaling, this indicates 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.

We therefore reconstruct the time series of latent satisfaction for Germany over the years 2017-2020 and display the result in figure 2. 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. In sum, the reconstructed chronicle of latent life satisfaction shows that 2020 was actually a year of lower subjective well-being.

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 yield similar results. Therefore, we shift our focus to France and use the CAMME dataset, a representative survey of the French population, conducted 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 of 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 of each year between 2016 and 2021. As the blue lines show, (i) there was no significant change in the nominal 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 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

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

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 latent life satisfaction in 2020.

4.2.3 The war in Ukraine

Although no socio-economic event can be univocally considered to have a negative effect on well-being, war is likely the closest to fitting this definition. We study the evolution of life satisfaction and rescaling during the ongoing war on Ukrainian territory.

We use the Ukrainian cross-section of the Gallup World Poll and 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, nominal levels of life satisfaction show no variation between 2018 and 2023 (blue lines). But the reconstructed M-LINE (in red) offers a different picture 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 has itself 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, the rescaling hypothesis suggests that the dip in latent satisfaction was deeper than what nominal satisfaction would indicate. In addition, these results provide 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) 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.

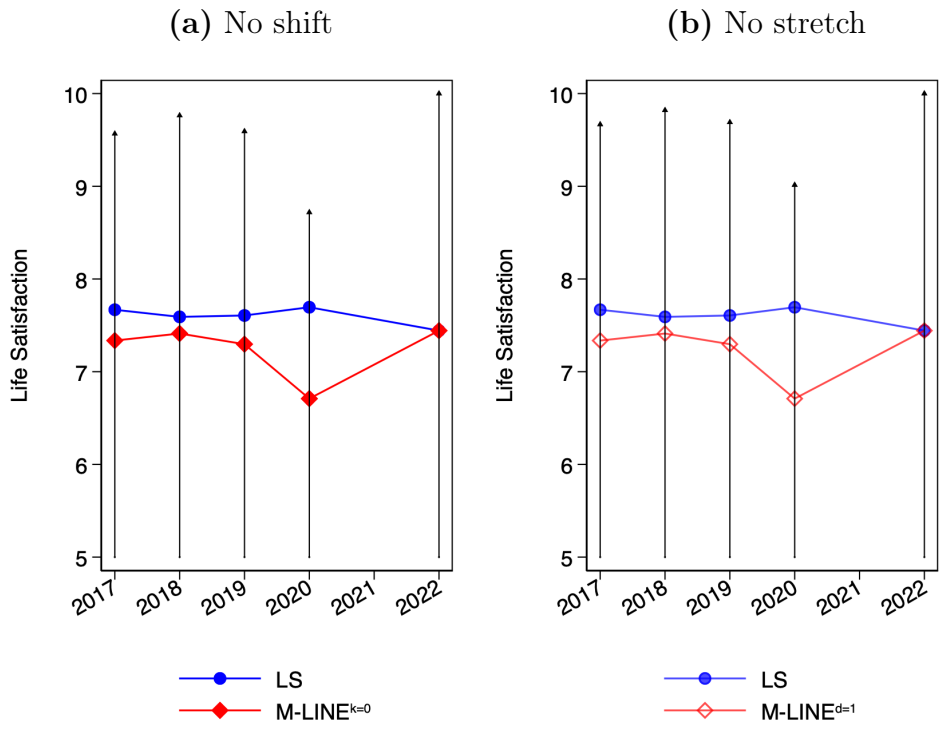


Figure 2: Well-being during the pandemic: M-LINE and nominal LS in Germany.

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]. $\bar{m} = 0.12$. Source SOEP-IS. Data is from different individuals interviewed in a nationally representative survey of the German population in 2022.

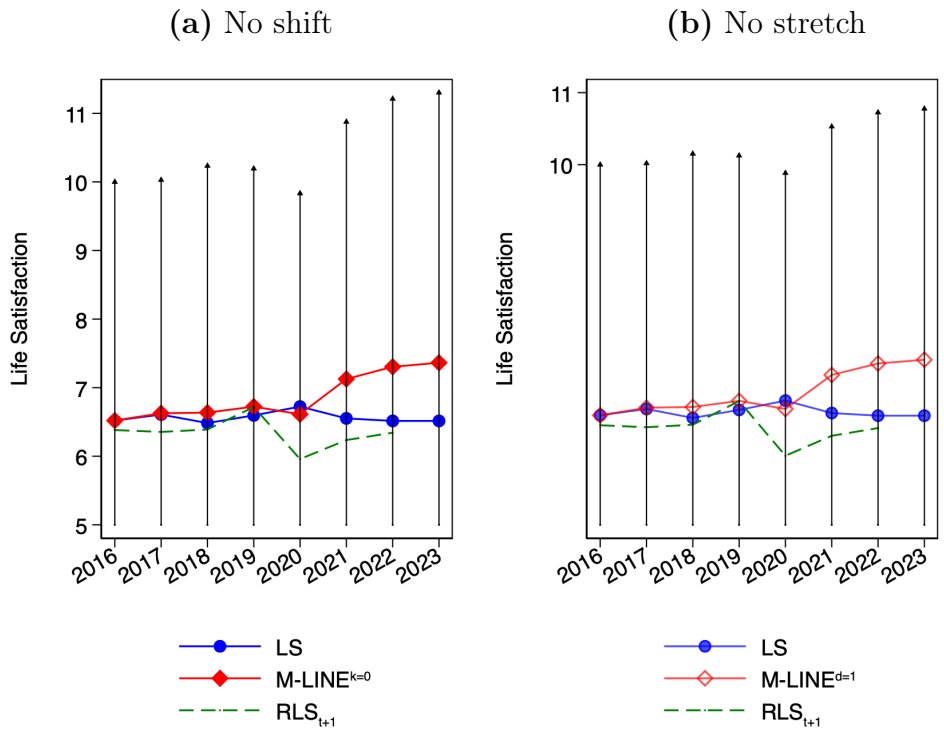


Figure 3: Well-being during the pandemic: M-LINE and nominal LS in France.

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). $\bar{m} = 0.12$. 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.

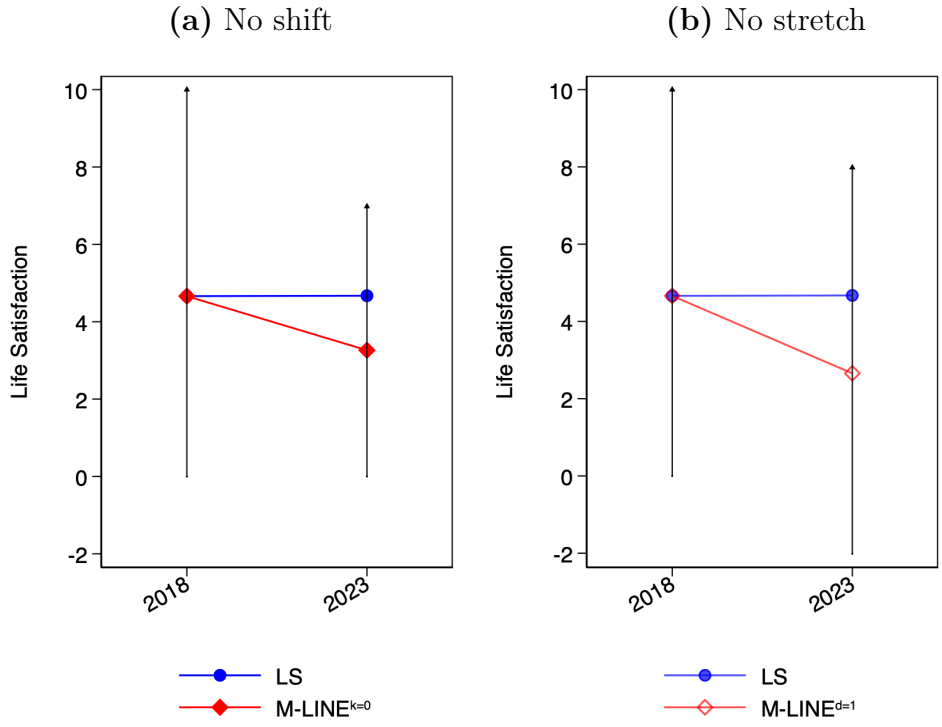


Figure 4: Well-being during the war: M-LINE and nominal LS in Ukraine.

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]. $\bar{m} = 0.12$. Source: Gallup World Poll. Data is from different individuals interviewed in two nationally representative surveys of the Ukrainian population in 2018 and 2023.

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. Additional analyses of reliability are reported in the appendix, section A7.

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 whether the M-LINE correlates with important life events or changes in people's objective life circumstances. This can be assessed by comparing the R squared in OLS regressions where life events are used to explain changes either in the M-LINE or in nominal life satisfaction LS. 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 (SOEP); (ii) an online panel of Prolific participants residing in the U.K. who were interviewed in 2023 and 2024.

Our first data source 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 a widower, moving to full-time unemployment ending unemployment, developing a new disability, and having a kid. Table 4 reports some summary statistics on the frequency of these events in our sample.

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

$$(h_{it} - h_{it-1}) = (X_{it} - X_{it-1})'\beta + (log(income)_{it} - log(income)_{it-1})'\gamma + (\epsilon_{it} - \epsilon_{it-1}) \tag{16}$$

where $h_{it} - h_{it-1}$ is a measure of well-being change (either ΔLS_{it} or ΔM-LINE$_{it}$), $(X_{it} - X_{it-1})$ 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 analyzes in the present section (table 5, columns (3) and (4); table 3, column (2)).[27]

[27]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). Admittedly, our choice of the trimming threshold is somewhat arbitrarily, and we must be sure that it does not mechanically generates any result. Therefore, we replicated all our regression analyses in Section 5 without any trimming (see table A7, table A11 and figure A15).

Table 3 reports the results of the regressions. The columns differ in the left-hand-side regressand, i.e., how changes in latent satisfaction are measured. Column (1) uses the standard 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 squared provide a direct measure of construct validity. They show that these events explain about 1% of the variations in nominal LS levels. This is similar to Kaiser and Prati (2025), who ran a similar exercise using 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 latent LS, but also the reporting LS scale in the same direction.

The estimated coefficients are mostly consistent across specifications and have the expected sign, although nominal LS seems to underestimate the positive short-term impact of marriage, and the M-LINE seems to underestimate the role of income changes. We do not have a satisfactory explanation for this latter result, which is somewhat surprising, and it is reversed in table 5, where we run the same analysis. The only important exception is childbirth. A long-standing puzzle in well-being research is why people deliberately have kids when this is detrimental to their self-reported 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.[28]

SOEP has some limitations as it only allows to study a limited set of events experienced by respondents (in the 1980s), but it is the only panel survey that has ever collected information on both retrospective and current life satisfaction. To complement this, we conduct our own data collection: a U.K. online panel (described in row 6 of table 2).

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 retrospective life satisfaction as of one year ago, again on a 0-100 scale, as well as detailed information about events that have happened in their lives. We adopt the exact same estimation strategy as with SOEP and use first-differences estimates. 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 with a new medical condition / incurring 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 analyses based on SOEP. The increased granularity of LS and RLS (now measured on a 100-point scale), together with the richer set of regressors, helps to increase 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

[28] 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.

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.

In the Appendix (tables A7-A14), we show that the M-LINE exhibits a consistently higher goodness of fit, regardless of our trimming strategy and across various calibrations of the memory parameter. In terms of regression coefficients, our results are robust to a $\bar{m}$ which is *smaller* than the estimated size. Imposing larger memory biases leads to a general attenuation of the estimated effects of positive events: the coefficients associated with income and marriage (in SOEP), with starting a new job and finishing off paying a debt (in the U.K. sample) all become null. The coefficient associated with childbirth is also attenuated, and may even flip sign.[29]

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

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

	(1) ΔLS	(2) ΔMLINE$^{k=0}$	(3) ΔMLINE$^{d=1}$
income change (log)	0.25***	0.15***	0.17***
	(0.04)	(0.05)	(0.03)
childbirth	-0.15***	0.20***	0.05
	(0.06)	(0.07)	(0.04)
marriage	0.21**	0.62***	0.39***
	(0.10)	(0.11)	(0.07)
end unemployment	0.92***	1.92***	1.04***
	(0.15)	(0.25)	(0.13)
start unemployment	-0.84***	-0.50***	-0.74***
	(0.09)	(0.10)	(0.07)
new disability	-0.14	0.09	-0.15*
	(0.12)	(0.14)	(0.08)
bereavement	-0.60**	-0.04	-0.80***
	(0.29)	(0.53)	(0.31)
N	27,485	27,485	27,485
R^2	0.011	0.015	0.022

Reading note: Linear regressions of change individual changes in life satisfaction on life events. Column (1) uses nominal life satisfaction; column (2) uses the M-LINE under the no-shift scenario; column (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]. $\bar{m} = 0.12$. Source: SOEP core.

[29] In every robustness test, we recalculate the trimming thresholds given the new $\bar{m}$, and rounding might lead to slightly different sample sizes (in the order of 0.1%-0.2%).

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

Category	**Nb of occurrences**	**N**
Marriage	434	27,485
Bereavement	70	27,485
End unemployment	283	27,485
Start unemployment	682	27,485
New disability	348	27,485
Childbirth	936	27,485
Income change*	24,363	27,485

* There were 24,363 episodes of annual household income changes. Income change is coded as Δ log(income). On average, the annual change was about +3%.

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 we observe someone's M-LINE increasing over a time period, we expect them to *declare* that they are more satisfied with their life today than before (and vice versa). In order to test this prediction, we collect data that allow measuring not only current and retrospective life satisfaction, but also the answer to the question "Compared to one year ago, how satisfied with your life are you nowadays?". The answer is measured on a continuous scale from -100 to +100 and labeled the "feeling better/worse" variable.[30]. We then estimate the correlation of this variable with the M-LINE.

We collected the necessary information from the U.K. online panel described above (N=990). Table 7 displays the correlation matrix of ΔLS, ΔM-LINE (no-stretch and no-shift scenarios) and the self-declared "feeling better/worse" variable. The estimated Pearson correlation between ΔLS and ΔM-LINE is $+0.39/+0.49$, 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 (+0.34), as expected. Crucially, the M-LINE has even stronger correlations with people's reports 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 lives (as witnessed by the correlation with objective variables), but also what is on people's minds (as shown by the correlation with the subjective feeling of doing better/worse).

[30]Recently, Kaiser and Prati (2025) studied the psychometric properties of comparative life satisfaction ratings and concluded that they outperform nominal life satisfaction levels on a variety of criteria. In the same vein, Köke and Perino (2017) ran a sort of horse-race between the variation of nominal LS over time and the direct assessment by respondents of whether their life had improved or deteriorated. They showed that the latter captures the changes in people's life better than the former.

Table 5: First-difference regression on life events (U.K., 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)	3.95**	4.17**	11.44***	11.31***	8.40***	8.21***
	(1.99)	(2.05)	(2.89)	(2.92)	(1.92)	(1.88)
childbirth	3.15	3.05	5.29**	4.40*	5.05***	4.31**
	(2.98)	(3.03)	(2.60)	(2.59)	(1.77)	(1.84)
new partner	2.33	4.61*	7.43**	7.82**	3.95	4.49
	(2.61)	(2.77)	(3.31)	(3.26)	(3.25)	(3.30)
fired	-3.03	-1.99	1.03	0.57	-2.19	-2.12
	(2.97)	(3.25)	(5.17)	(5.18)	(3.17)	(3.25)
quit job	-0.94	-0.49	2.28	1.83	2.51	2.42
	(2.41)	(2.43)	(3.05)	(3.11)	(2.31)	(2.38)
start debt	-3.27**	-2.59	-1.10	-1.94	-1.93	-2.32*
	(1.58)	(1.58)	(1.55)	(1.57)	(1.32)	(1.35)
end debt	0.94	1.01	3.92***	3.22**	3.36***	2.77***
	(1.12)	(1.16)	(1.22)	(1.28)	(1.01)	(1.04)
new diagnosis/injury	-1.66	-0.05	-0.68	-0.62	-1.04	-0.91
	(1.15)	(1.33)	(1.09)	(1.20)	(0.94)	(1.07)
new job	3.41*	3.47**	5.82**	4.76*	3.28**	2.64*
	(1.76)	(1.75)	(2.88)	(2.71)	(1.64)	(1.58)
promoted	-0.51	-0.47	2.82*	2.33	3.13**	2.93**
	(1.52)	(1.59)	(1.67)	(1.78)	(1.36)	(1.44)
bereavement/loss	-1.86*	-1.61	0.61	-0.19	-0.42	-0.79
	(1.09)	(1.23)	(1.02)	(1.14)	(0.90)	(0.94)
Additional events		✓		✓		✓
N	990	990	990	990	990	990
R^2	0.028	0.060	0.112	0.140	0.100	0.124

Reading note: Linear regressions of 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]. $\bar{m} = 1.2$ on a 0-100 scale. 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. Source: U.K. online panel.

Table 6: Recorded events (U.K., 2023-24)

Category	**Nb of occurrences**	**N**
Childbirth	28	990
New Partner	36	990
Fired	24	990
Quit Job	68	990
Start Debt	133	990
End Debt	180	990
New Diagnosis/Injury	170	990
New Job	138	990
Promoted	90	990
Bereavement/Loss	259	990
Income change*	649	990

* 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

Reading note: Pearson correlation coefficients. * $p < 0.01$. 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]. Feeling better/worse: "Compared to one year ago, how satisfied with your life are you nowadays?". $\bar{m} = 1.2$ on a 0-100 scale. Source: U.K. online panel.

5.3 Predictive validity: does the M-LINE predict important life decisions?

The ability to predict future behaviors is a fundamental benchmark for assessing the quality of a subjective measure. Some previous contributions (Freeman, 1978; Clark, 2001; Frisch et al., 2005; Kaiser and Oswald, 2022b) have shown that satisfaction data in a given year t can predict subsequent decisions one year later. Here, we run a similar exercise and compare the relative predictive power of ΔLS and ΔM-LINE. This can be assessed by comparing the goodness-of-fit in an OLS regression where the dependent variable is a life event triggered by individual decisions. To run this exercise, we need a panel data set that contains at least three years of observations (two initial years for the satisfaction data, and one later year for the decisions to be predicted), as well as a sufficient number of decisions to be predicted. We therefore use the German SOEP, and attempt to predict some plausibly endogenous events from the regressors' list in table 3, namely: getting married, the birth of a child, and ending unemployment. We use a linear estimator with standard errors clustered at the individual level and regress a one-year-ahead binary variable ($event_{t+1}$) on a well-being change measure (Δh_t).

Figure 5 displays the standardized OLS coefficients - which, in a single-regressor model, are a direct measure of goodness-of-fit. Reading from the top to the bottom, the dependent variables are $\Delta\text{MLINE}^{\text{k}=0}$, $\Delta\text{MLINE}^{\text{d}=1}$ and ΔLS. The three panels correspond to the three events we are trying to predict. We base our analysis on almost 30,000 individual-year responses, running from 1985 to 1988.

To start, let us discuss the main similarity, i.e., that the coefficients associated with the M-LINE and ΔLS are consistent in sign. The positive coefficients associated with marriage and childbirth indicate that an increase in life satisfaction is associated with a slightly higher probability of getting married or having a child in the coming year, probably reflecting a positive trend in a romantic relationship. Instead, those who endure a deterioration in their life satisfaction are more likely to end their unemployment spell in the coming year, in an exit type of behavior (as in Mavridis, 2015).

While the signs of the estimated coefficients are similar, the significance and size not always are. We can reject the null hypothesis that the coefficient on our predictor is zero for all M-LINE coefficients, while ΔLS is not a significant predictor of marriage. The two panels on the left-hand side show that, compared to nominal LS, one standard deviation change in the M-LINE is associated with a greater increase in the probability of getting married or having a child. The associated R squared (not displayed in the image) are higher too. The third panel shows that, when predicting the end of unemployment, ΔM-LINE and ΔLS give pretty similar estimates. Importantly, the M-LINE also provides significant *incremental* predictive validity. That is, it provides useful additional information in predicting childbirth and marriage in a model where both ΔLS and ΔM-LINE are used as predictors (see the Appendix, figure A11).

In the Appendix, we run some additional analyses, and show that the superior predictive validity of the M-LINE for marriage and childbirth is confirmed when using different values of $\bar{m}$ or when using no trimming (figures A12-A14). As a final check, we run a placebo test, where we attempt

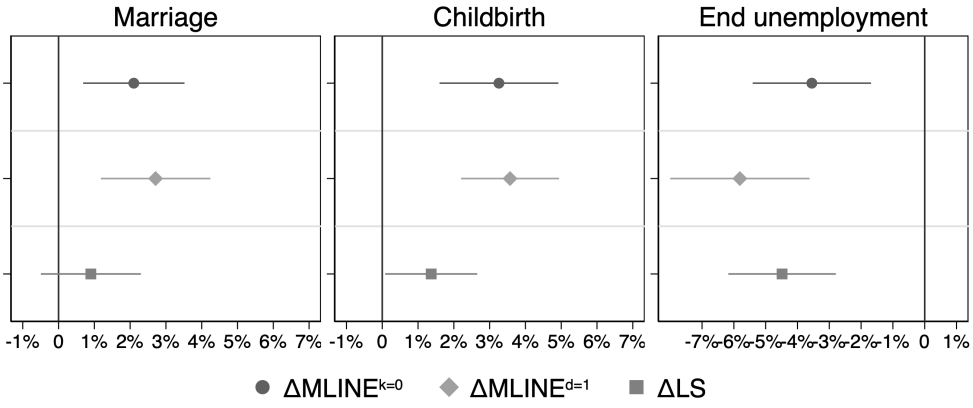


Figure 5: Probability of a future event (SOEP, 1985-1988)

Reading note: Standardized OLS coefficients from the regression of an event on the lagged measure of well-being change. LS question: "How satisfied are you with your life, all things considered?" [0-10]. RLS question: "How satisfied were you a year ago with your life?" [0-10]. $\bar{m} = 0.12$. The dataset contains 595 weddings, 1,290 childbirths and 363 end of unemployment spells. Source: SOEP core.

to predict a plausibly exogenous and unforseeable event, i.e., moving out because of a contract termination by the landlord. This is a plausibly exogenous event (it's a decision of the landlord, not of the tenant) and plausibly unforeseeable one year ahead (in Germany, the termination notice is typically given 3-6 months in advance). When we run the prediction exercise on this event, neither LS nor the M-LINE display any prediction power, as expected (figure A16).

Overall, these results show that the M-LINE tends to be better than (for family decisions) or as good as (for job market decisions) ΔLS at predicting future behaviors. Once again, this is indicative of the fact that using the M-LINE may help capture part of the information that is otherwise hidden by the rescaling process.

5.4 Subjective perceptions of rescaling

Is rescaling an abstract "ivory tower" concept, or is it something people actually perceive? To inquire, we included some questions in our U.S. online panel survey (N=1,627). 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 (42%) 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 32% perceived it as lower today and 18% as higher than 5 years ago. Table A6 in the Appendix, displays the whole cross-tabulation. 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.

These results show that a sizeable share (30%) of respondents do not perceive any change in their scale over five years. When we look at the association between perceived rescaling and basic demographic characteristics, we find that older adults are less likely to rescale than younger ones (see Appendix, table A22). The negative association between age and rescaling also appears when we regress estimated (vs perceived) rescaling parameters on a rich array of sociodemographic variables using the U.K. online panel (see Appendix, tables A23 and A24). Overall, shifting and stretching seem to become less common with older age, in particular upward movements. This has potentially important implications for how we assess well-being changes over the life course - a research avenue which is certainly worth exploring. It also warrants caution when correcting for rescaling in countries that have very different demographic structures, since some age-related adjustment might be needed.

5.5 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) retrospective life satisfaction; (ii) the memory bias parameter. 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 as of 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. We do *not* recommend using the $\bar{m} = 0.12$ for non-Western countries, since the memory bias could be sensitive to cultural and demographic factors that cannot be plausibly assumed to be constant with respect to our calibration study. Below, we provide some concrete instructions for data collectors.

Data structure. The M-LINE can be calculated either from cross-sectional data or from panel data. While the former allows to make corrections for rescaling at the aggregate (country) level, panel data allows to calculate individual-level rescaling parameters. We remind practitioners that

for micro-regressions using panel data, they must use a scale with a lower bound of 1 (*not* 0) and we recommend employing some trimming that discards implausible rescaling. For more details, we invite readers to review section 5.1.

Frequency of data collection. Retrospective 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 as of one year ago. This is the best solution in a repeated cross-section as well. The recommendations below refer to surveys that have at least one wave per year.

Questions and wording Ideally, every year, around the same period, one survey wave should contain: (1) the standard life satisfaction question (for instance, the one from the U.K. Office for National Statistics: "Overall, how satisfied are you with your life nowadays?") as well as: (2) a question on retrospective life satisfaction ("And, overall, how satisfied were you with your life about one year ago?"). On top of that, at least for the first wave of data collection, it is important to insert: (3) a question on retrospective satisfaction two weeks ago ("And, overall, how satisfied were you with your life about two weeks ago?").

Order of the questions As is often recommended, the life satisfaction question LSIQ1 should be asked first, so it is not contaminated by the following questions. Ideally, retrospective life satisfaction RLS should follow immediately.

Logic The mean answer to question (2) in year t can be compared to the mean answer to question (1) in year $t-1$. The discrepancy between the two contains information about rescaling. The answer to question (3) can be compared to the answer to question (1). The discrepancy between the two contains information about memory biases.

Implementation Questions (1) and (2) must be in the same survey wave. Similarly, questions (1) and (3) must be in the same wave. However, there is no need for (2) and (3) to be in the same wave. Question (2) must always be asked around the same period, i.e., at 12-month distance. This requirement is less stringent for question (3). Question (2) must be asked each year. Question (3) can be asked much less frequently, or only once.

6 Conclusion

We have explored the consequences of interpreting the two Easterlin paradoxes in terms of rescaling rather than hedonic treadmill. Our method is based on the exploitation of current and retrospective self-reported scores of life satisfaction. We interpret the difference between the level of past satisfaction as declared at 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. This requires giving some credit to respondents' self-reported current and retrospective life satisfaction and not attributing potential gaps between these levels entirely to recall biases.

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. People also seem to be aware about their own changing interpretation of the LS scale over time. 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 of raising national happiness in a durable way (and little risk of reducing it). If the M-LINE is used instead, national happiness will be more sensitive to what happens in a society, such as changes in income, health, education, pandemics, and wars.

We do not claim to have demonstrated that rescaling, rather than the 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, excluding one interpretation or the other is simply out of reach. There is also little reason to consider the two phenomena mutually exclusive. 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 USA 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 USA 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 to attenuate the Easterlin paradox.

Our exercise has some limitations. First, we are able to reconstruct a long-run 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 hope that the method consisting of using both 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.

Alesina, A., Di Tella, R., and MacCulloch, R. (2004). Inequality and Happiness: Are Europeans and Americans Different? *Journal of Public Economics*, 88(9-10):2009–2042.

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.

Anvari, F. (2023). *Unpublished dataset*.

Barrett, L. F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. *Social cognitive and affective neuroscience*, 12(1):1–23.

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.

Benjamin, D. J., Heffetz, O., Kimball, M. S., and Rees-Jones, A. (2012). What do you think would make you happier? What do you think you would choose? *American Economic Review*, 102(5):2083–2110.

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.

Case, A. and Deaton, A. (2020). *Deaths of Despair and the Future of Capitalism*. Princeton University Press, Princeton, NJ.

Clark, A. (2016). SWB as a Measure of Individual Well-Being. In Adler, M. and Fleurbaey, M., editors, *The Oxford Handbook of Well-Being and Public Policy*.

Clark, A. E. (2001). What really matters in a job? Hedonic measurement using quit data. *Labour economics*, 8(2):223–242.

Clark, A. E., Frijters, P., and Shields, M. A. (2008). Relative income, happiness, and utility: An explanation for the Easterlin paradox and other puzzles. *Journal of Economic Literature*, 46(1):95–144.

Clark, A. E. and Senik, C. (2010). Who compares to whom? The anatomy of income comparisons in Europe. *The Economic Journal*, 120(544):573–594.

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.

De Neve, J.-E., Ward, G., De Keulenaer, F., Van Landeghem, B., Kavetsos, G., and Norton, M. I. (2018). The asymmetric experience of positive and negative economic growth: Global evidence using subjective well-being data. *The Review of Economics and Statistics*, 100(2):362–375.

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.

Di Tella, R. and MacCulloch, R. (2010). Happiness adaptation to income beyond 'basic needs'. In Diener, E., Helliwell, J., and Kahneman, D., editors, *International Differences in Well-Being*, chapter 8, pages 217–247. Oxford University Press, New York.

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

Easterlin, R. A. (1974). Does economic growth improve the human lot? Some empirical evidence. In *Nations and households in economic growth*, pages 89–125. Academic Press.

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., McVey, L. A., Switek, M., Sawangfa, O., and Zweig, J. S. (2010). The happiness-income paradox revisited. *Proceedings of the National Academy of Sciences*, 107(52):22463–22468.

Easterlin, R. A. and O'Connor, K. (2020). The Easterlin Paradox. IZA Discussion Paper 13923, Institute of Labor Economics (IZA).

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*.

Freeman, R. B. (1978). Job satisfaction as an economic variable. *American Economic Review*, 68:135–141.

Frisch, M. B., Clark, M. P., Rouse, S. V., Rudd, M. D., Paweleck, J. K., Greenstone, A., and Kopplin, D. A. (2005). Predictive and treatment validity of life satisfaction and the quality of life inventory. *Assessment*, 12(1):66–78.

Goff, L. (2024). Identifying causal effects with subjective ordinal outcomes. *arXiv:2212.14622*.

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., De Neve, J.-E., Aknin, L. B., and Wang, S., editors (2025). *World Happiness Report 2025*. Wellbeing Research Centre, University of Oxford, Oxford, UK.

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*.

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 Deaton, A. (2010). High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Sciences*, 107(38):16489–16493.

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., Krueger, A. B., Schkade, D., Schwarz, N., and Stone, A. A. (2006). Would You Be Happier If You Were Richer? A Focusing Illusion. *Science*, 312(5782):1908–1910.

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

Kahneman, D., Wakker, P. P., and Sarin, R. (1997). Back to Bentham? Explorations of experienced utility. *The quarterly journal of economics*, 112(2):375–406.

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

Kaiser, C. and Lepinteur, A. (2025). Measuring the unmeasurable? systematic evidence on scale transformations in subjective survey data. *University of Oxford Wellbeing Research Centre Working Paper 2501*.

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. (2025). Why you should measure subjective changes. *University of Oxford Wellbeing Research Centre Working Paper 2501*.

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

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.

Kleckner, I. R., Zhang, J., Touroutoglou, A., Chanes, L., Xia, C., Simmons, W. K., Quigley, K. S., Dickerson, B. C., and Feldman Barrett, L. (2017). Evidence for a large-scale brain system supporting allostasis and interoception in humans. *Nature Human Behaviour*, 1(5):0069.

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.

Layard, R. (2005). *Happiness: Lessons from a New Science*. Penguin Press, New York.

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.

Mavridis, D. (2015). The unhappily unemployed return to work faster. *IZA Journal of Labor Economics*, 4(2).

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.

Oishi, S., Kesebir, S., and Diener, E. (2011). Income inequality and happiness. *Psychological Science*, 22(9):1095–1100.

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.

Sacks, D. W., Stevenson, B., and Wolfers, J. (2012). The new stylized facts about income and subjective well-being. *Emotion*, 12(6):1181–1187.

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`.

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.

Tobler, P. N., Fiorillo, C. D., and Schultz, W. (2005). Adaptive coding of reward value by dopamine neurons. *Science*, 307(5715):1642–1645.

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

Volkow, N. D., Fowler, J. S., Wang, G.-J., Hitzemann, R., Logan, J., Schlyer, D. J., Dewey, S. L., and Wolf, A. P. (1993). Decreased dopamine D2 receptor availability is associated with reduced frontal metabolism in cocaine abusers. *Synapse*, 14(2):169–177.

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.

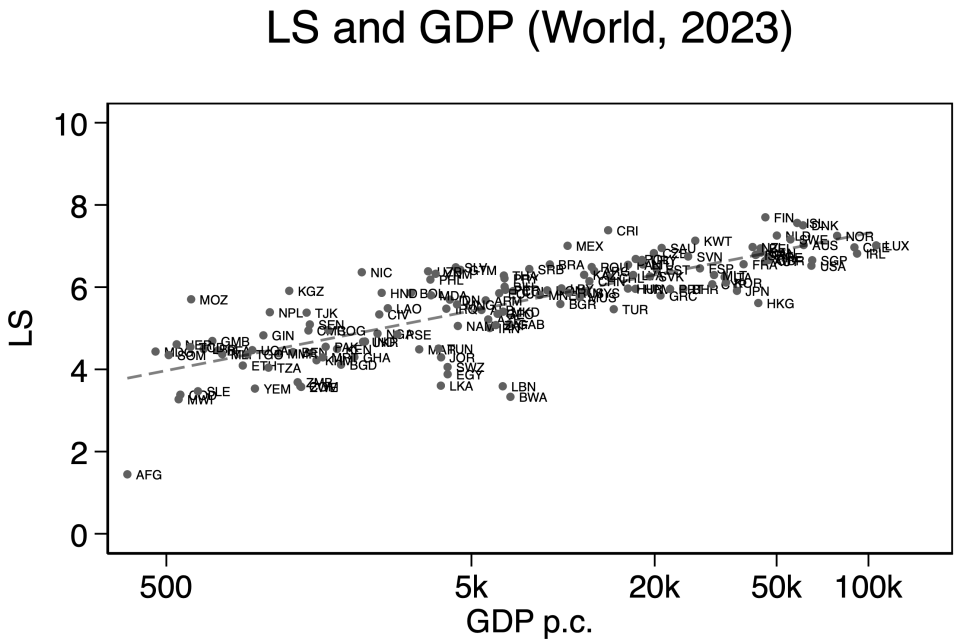
Acknowledgments

We are grateful for financial support from the French National Research Agency, through the Investissements d'avenir framework ANR-17-EURE-0001; from the Institut Universitaire de France; 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, Olivier Blanchard, 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, Bernard Van Praag. 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?
ONLINE APPENDIX

Alberto Prati Claudia Senik

A1 Additional Figures



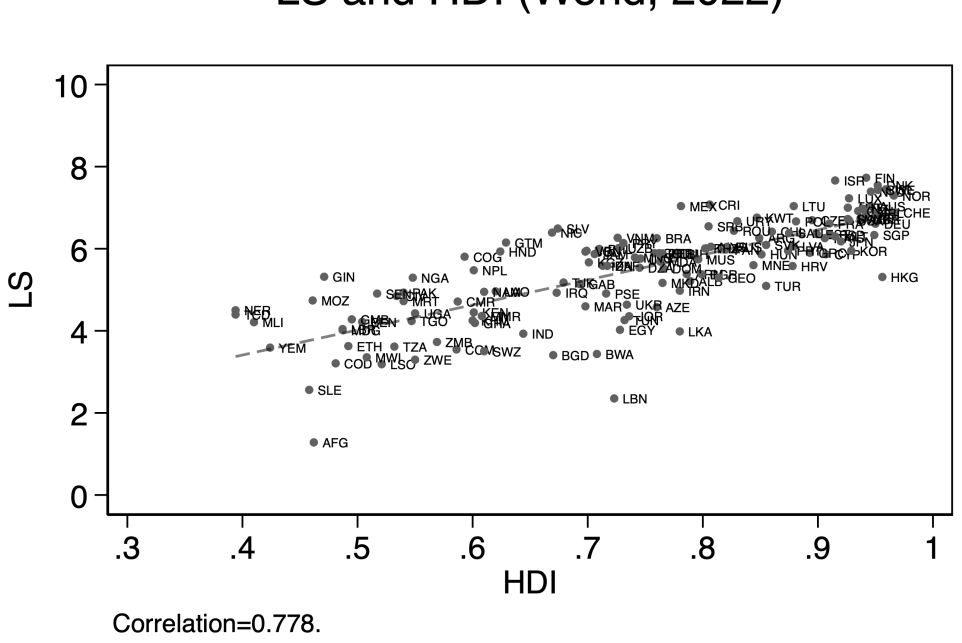
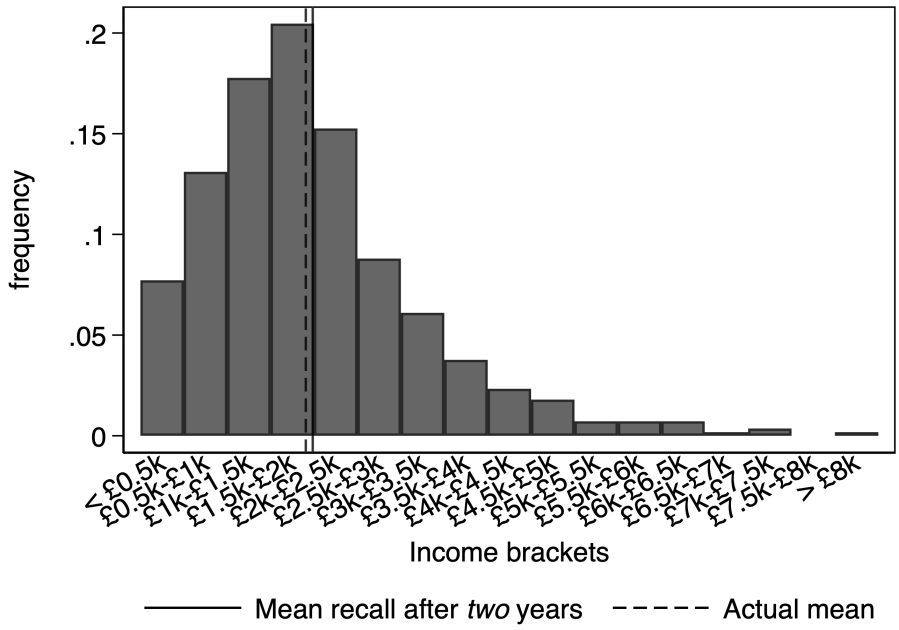
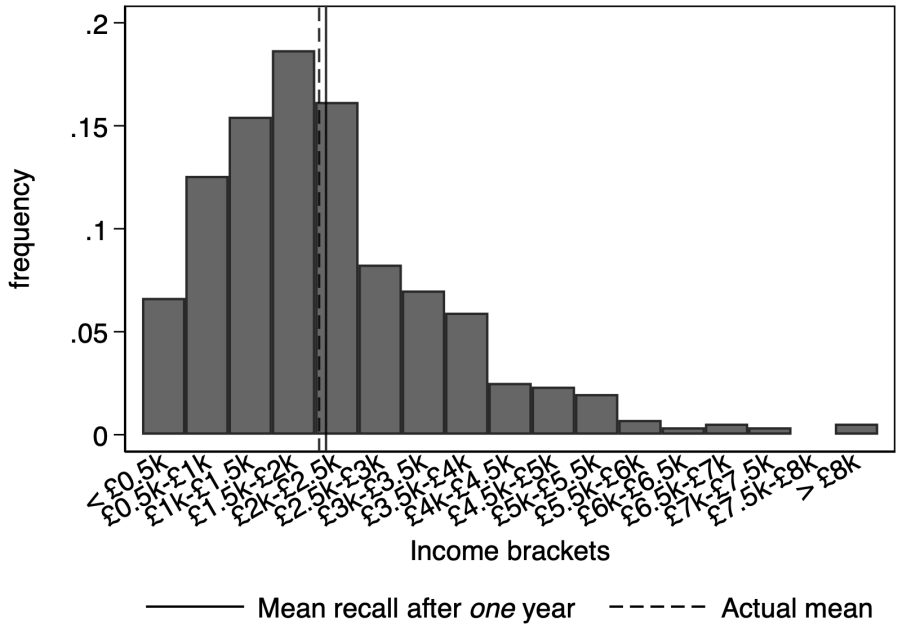


Figure A1: LS and objective well-being (cross-section).

Reading note: At a point in time, richer and more developed countries are also happier. $N = 139$ countries (GDP) and 138 countries (HDI). 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.



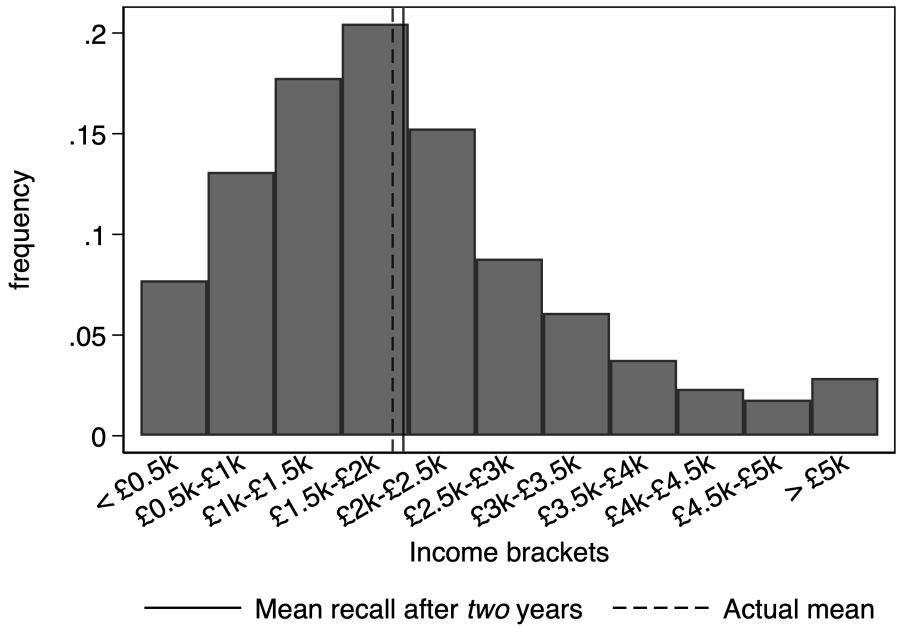
(a) 2023 monthly income (as recalled in 2025)



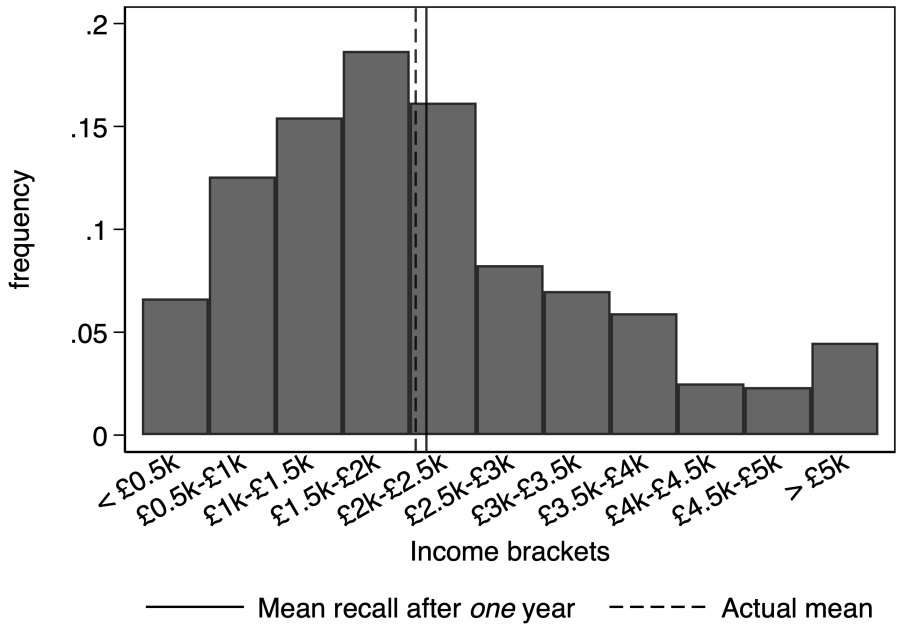
(b) 2024 monthly income (as recalled in 2025)

Figure A2: Memory bias for income

Reading note: Panel (a) shows the distribution of 2023 income, as recalled in 2025. Panel (b) shows the distribution of 2024 income, as recalled in 2025. In each panel, the vertical dotted bar shows the actual mean and the dotted bar the recalled mean. The difference between the two is the recall bias. Income question: "What is your monthly personal income before tax (to the closest £500)?" [asked in 2023, 2024, 2025]; "And what was your monthly personal income before tax 12 months ago?" [asked in 2025]; "And what was your monthly personal income before tax two years ago?" [asked in 2025]. The sample is a balanced longitudinal panel of 557 respondents who were interviewed in February-March 2023, 2024 and 2025. Data source: UK online panel.



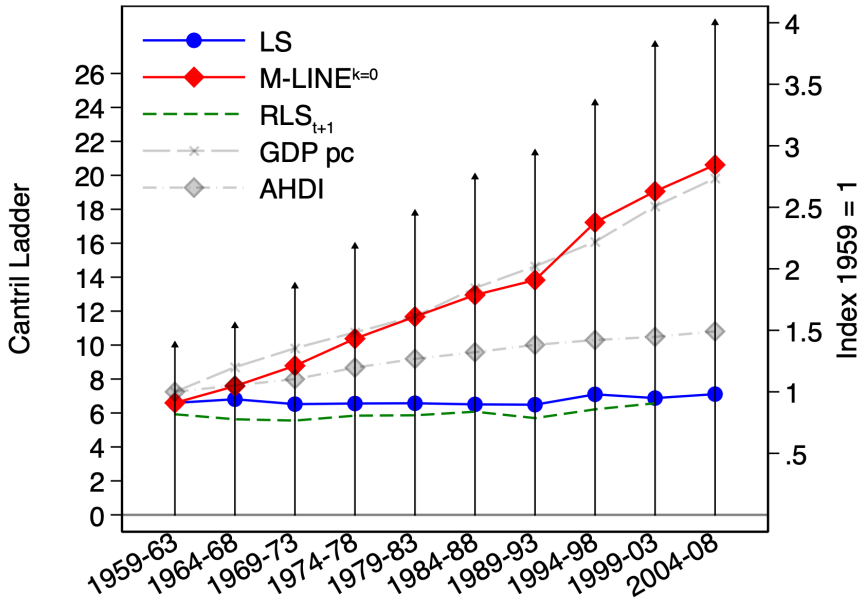
(a) 2023 monthly income (as recalled in 2025)



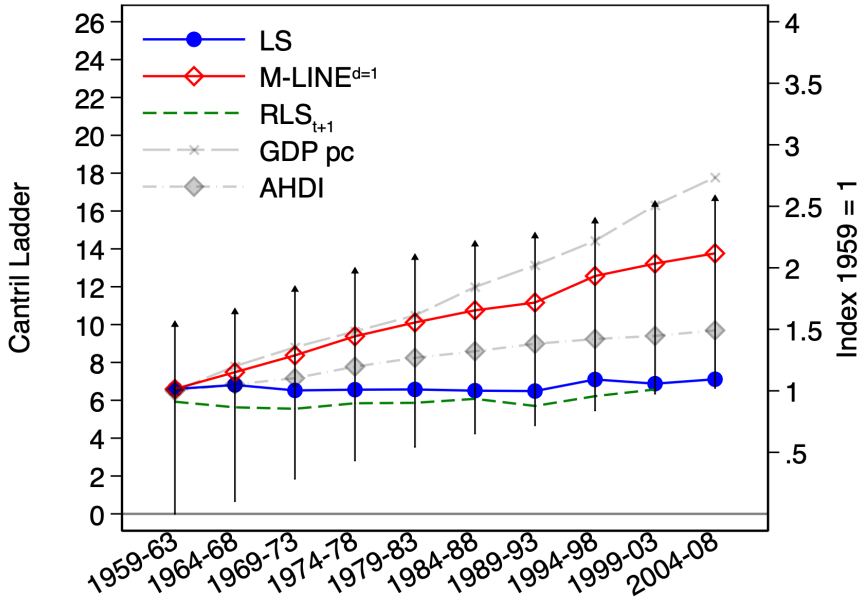
(b) 2024 monthly income (as recalled in 2025)

Figure A3: Memory bias for income (11 categories)

Reading note: Panel (a) shows the distribution of 2023 income, as recalled in 2025. Panel (b) shows the distribution of 2024 income, as recalled in 2025. Contrary to Fig. A2, here the income brackets are top-coded, so that the income scale has 11 categories instead of sixteen. In each panel, the vertical dotted bar shows the actual mean and the dotted bar the recalled mean. The difference between the two is the recall bias. Income question: "What is your monthly personal income before tax (to the closest £500)?" [asked in 2023, 2024, 2025]; "And what was your monthly personal income before tax 12 months ago?" [asked in 2025]; "And what was your monthly personal income before tax two years ago?" [asked in 2025]. The sample is a balanced longitudinal panel of 557 respondents who were interviewed in February-March 2023, 2024 and 2025. Data source: U.K. online panel.

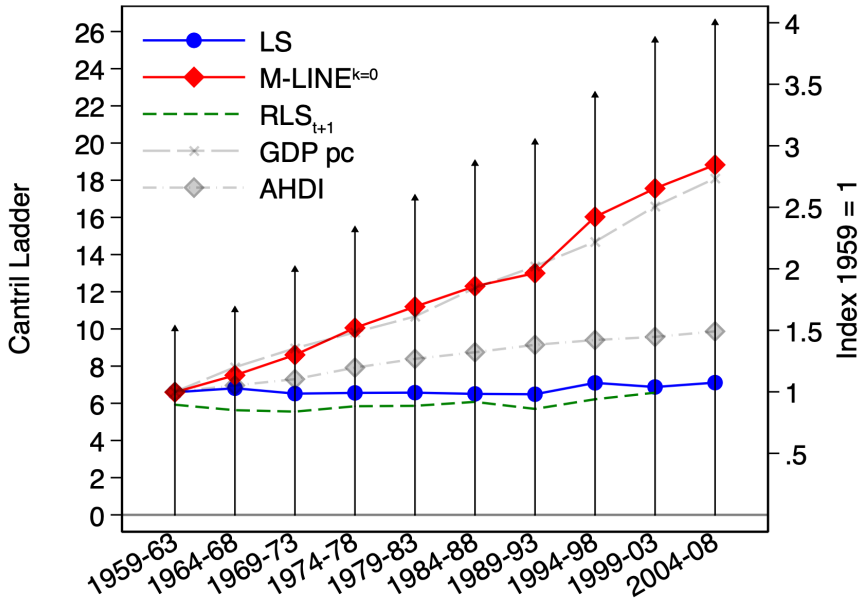


(a) No shift ($\bar{m}$=**0.00**)

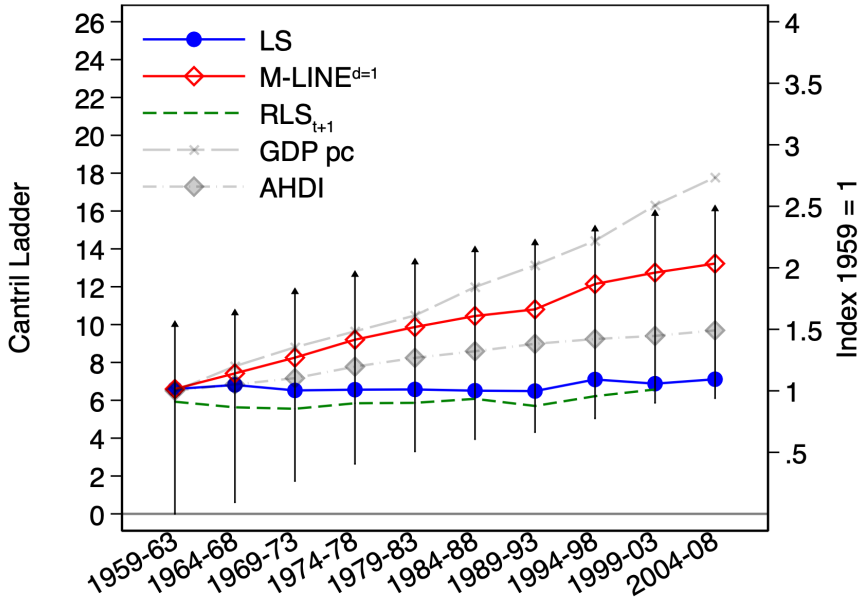


(b) No stretch ($\bar{m}$=**0.00**)

Figure A4: Latent satisfaction for different rescaling patterns (USA, 1959-2008) for $\bar{m}$=**0.00**

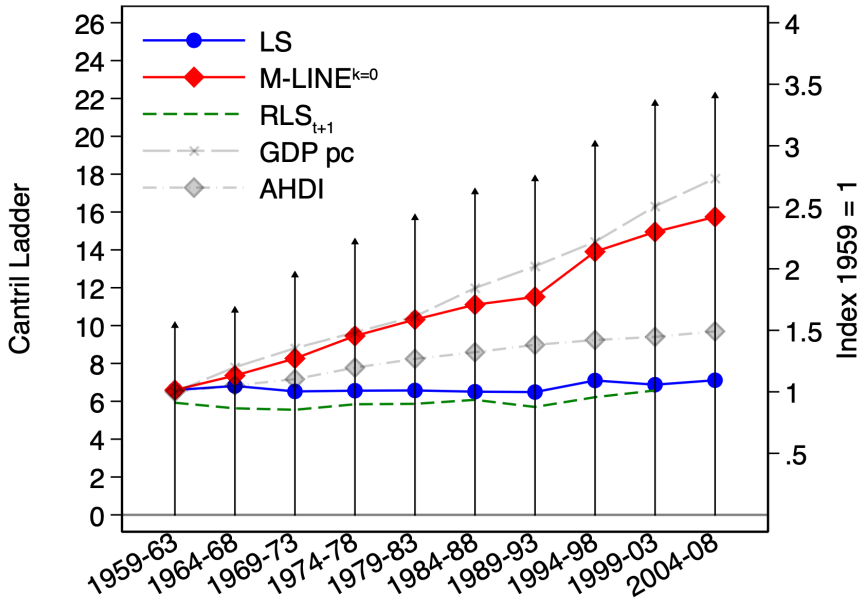


(a) No shift ($\bar{m}$=**0.06**)

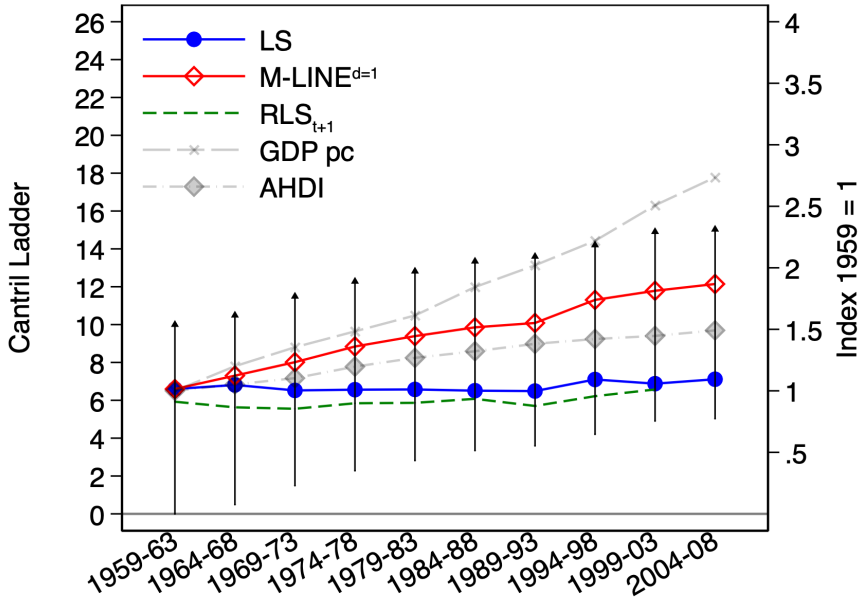


(b) No stretch ($\bar{m}$=**0.06**)

Figure A5: Latent satisfaction for different rescaling patterns (USA, 1959-2008) for $\bar{m}$=**0.06**

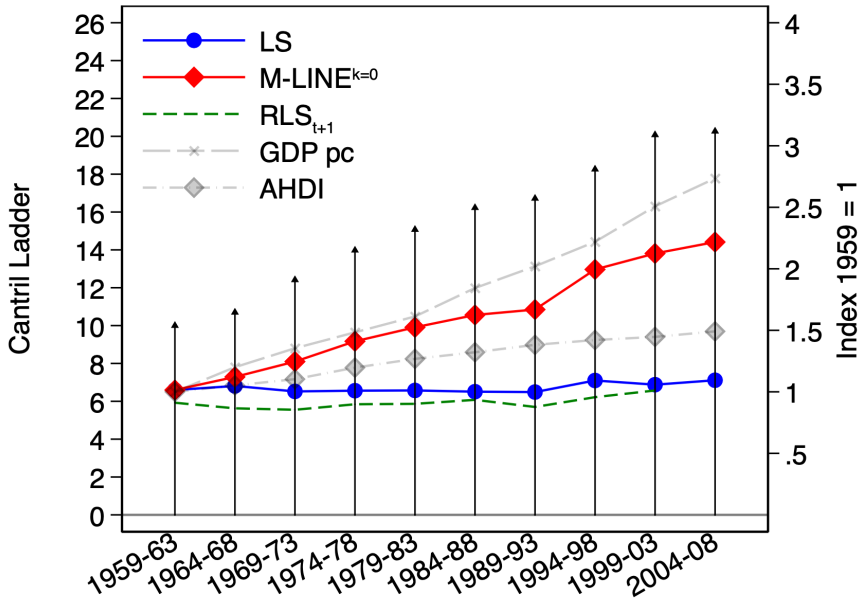


(a) No shift ($\bar{m}$=**0.18**)

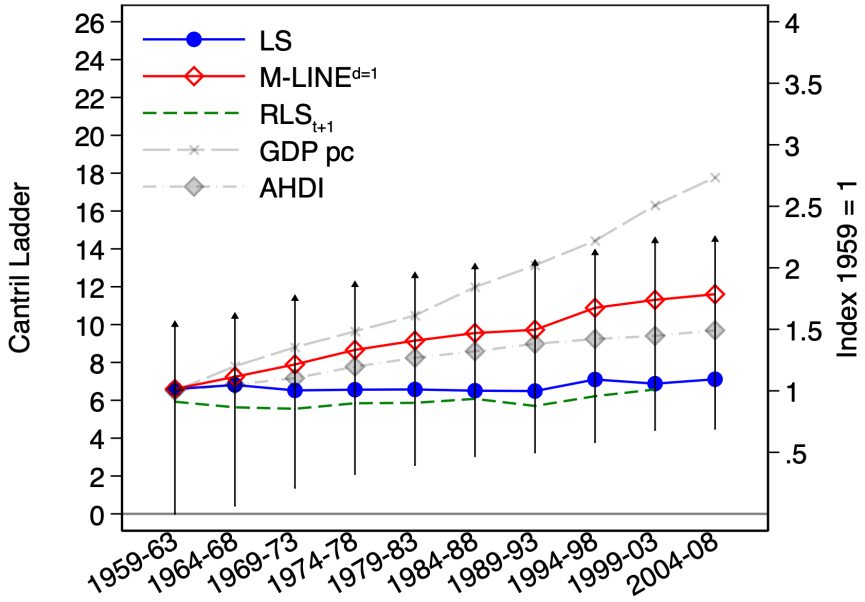


(b) No stretch ($\bar{m}$=**0.18**)

Figure A6: Latent satisfaction for different rescaling patterns (USA, 1959-2008) for $\bar{m}$=**0.18**

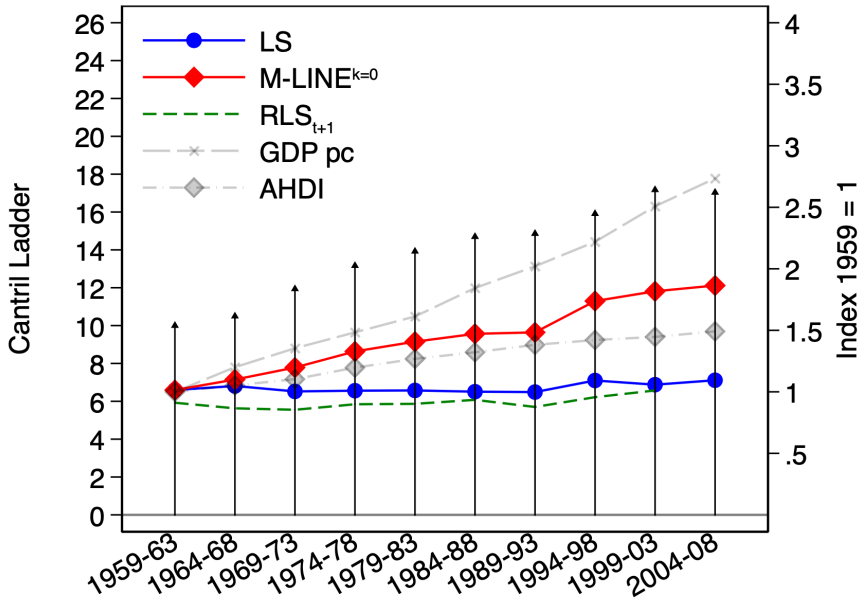


(a) No shift ($\bar{m}$=**0.24**)

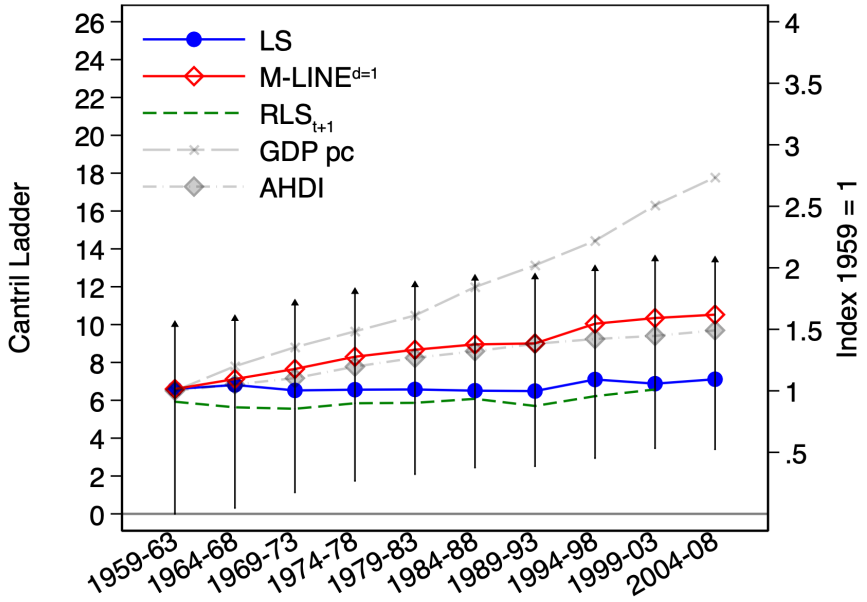


(b) No stretch ($\bar{m}$=**0.24**)

Figure A7: Latent satisfaction for different rescaling patterns (USA, 1959-2008) for $\bar{m}$=**0.24**

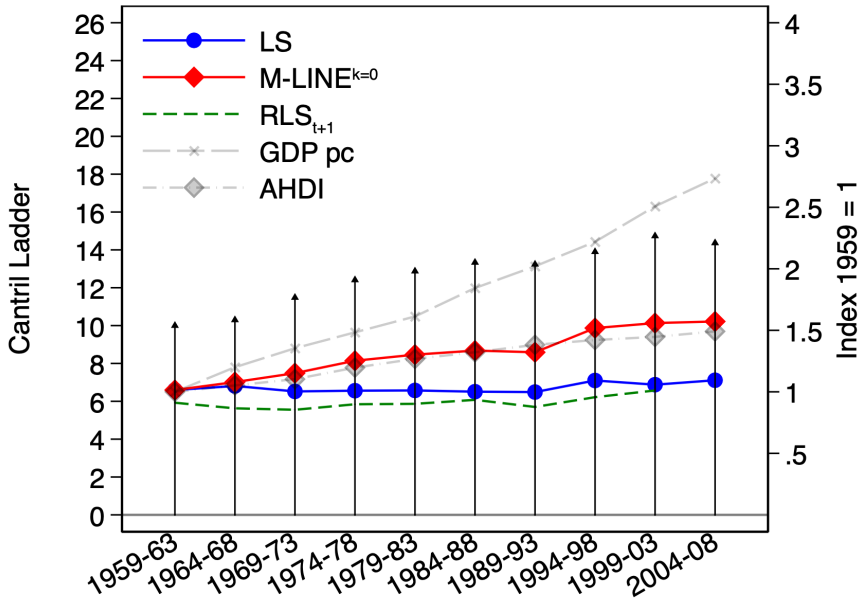


(a) No shift ($\bar{m}$=**0.36**)

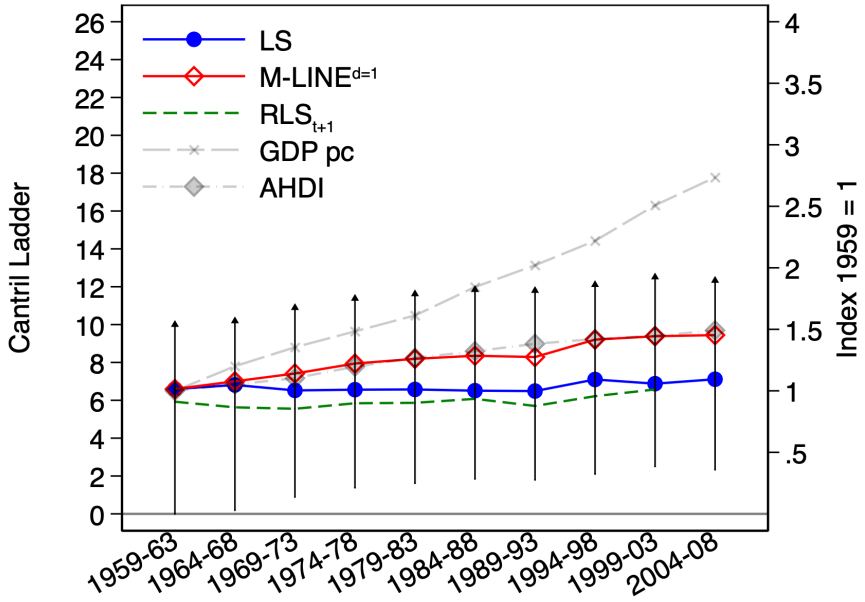


(b) No stretch ($\bar{m}$=**0.36**)

Figure A8: Latent satisfaction for different rescaling patterns (USA, 1959-2008) for $\bar{m}$=**0.36**

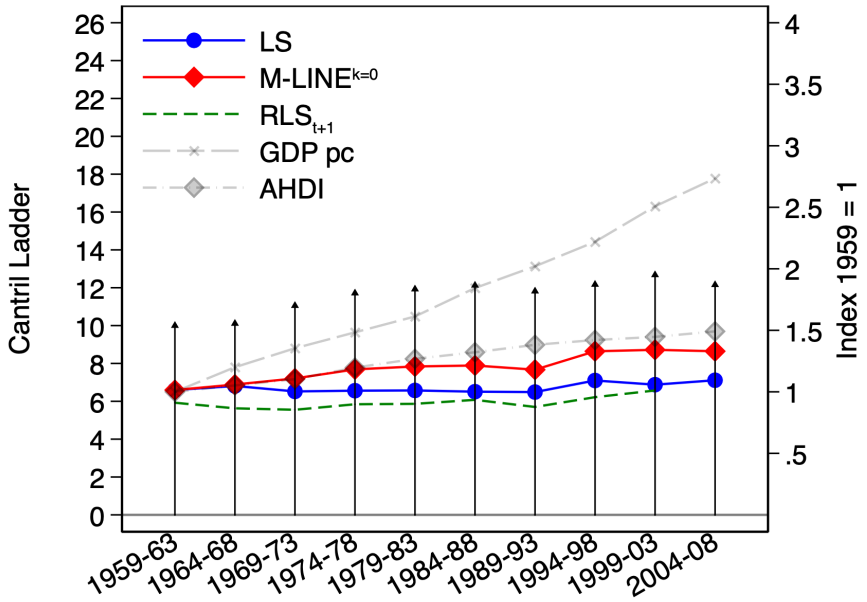


(a) No shift ($\bar{m}$=**0.48**)

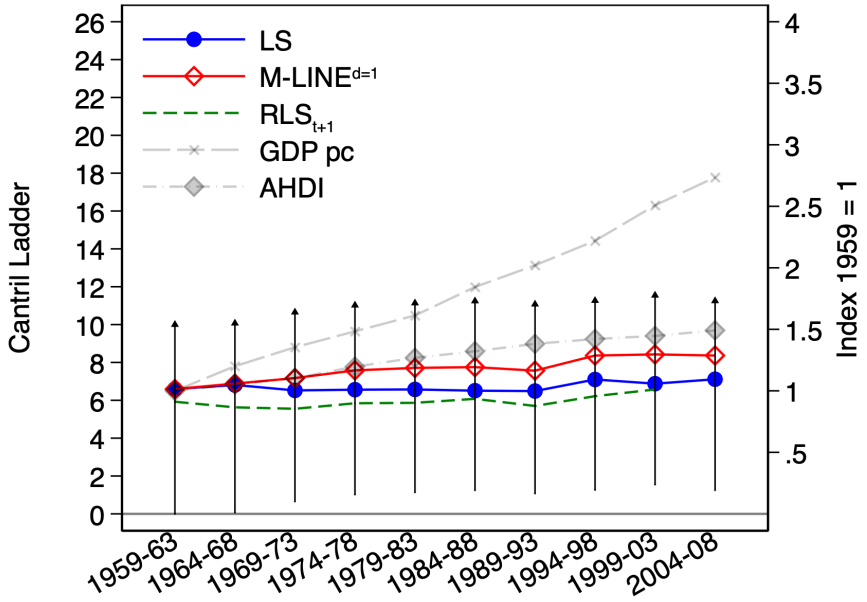


(b) No stretch ($\bar{m}$=**0.48**)

Figure A9: Latent satisfaction for different rescaling patterns (USA, 1959-2008) for $\bar{m}$=**0.48**



(a) No shift ($\bar{m}$=**0.60**)



(b) No stretch ($\bar{m}$=**0.60**)

Figure A10: Latent satisfaction for different rescaling patterns (USA, 1959-2008) for $\bar{m}$=**0.60**

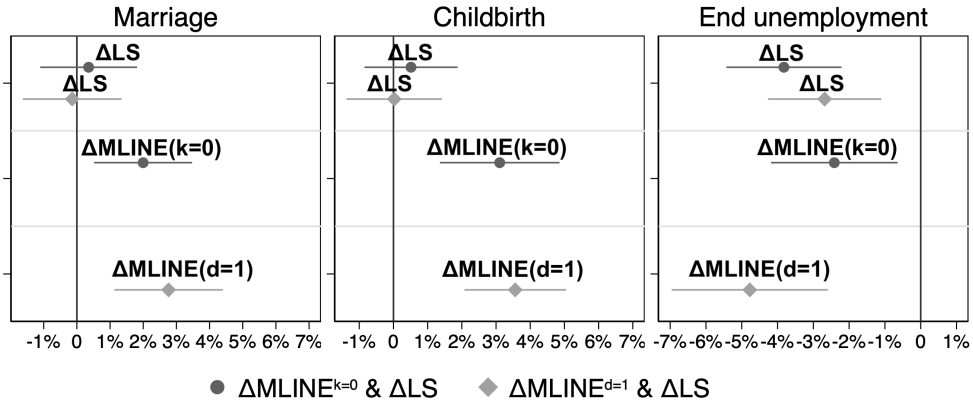


Figure A11: Probability of a future event when both ΔLS and Δ M-LINE are included as right-hand-side variables

Reading note: Standardized OLS coefficients from the regression of an event on the lagged measure of well-being change. LS question: "How satisfied are you with your life, all things considered?" [0-10]. RLS question: "How satisfied were you a year ago with your life?" [0-10]. Source SOEP.

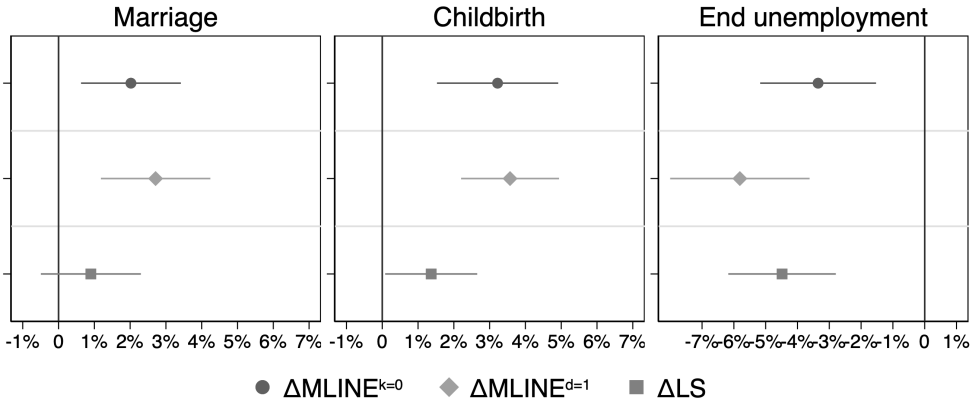


Figure A12: Robustness check: figure 5 with $\bar{m} = 0$

Reading note: Standardized OLS coefficients from the regression of an event on the lagged measure of well-being change. LS question: "How satisfied are you with your life, all things considered?" [0-10]. RLS question: "How satisfied were you a year ago with your life?" [0-10]. Source SOEP.

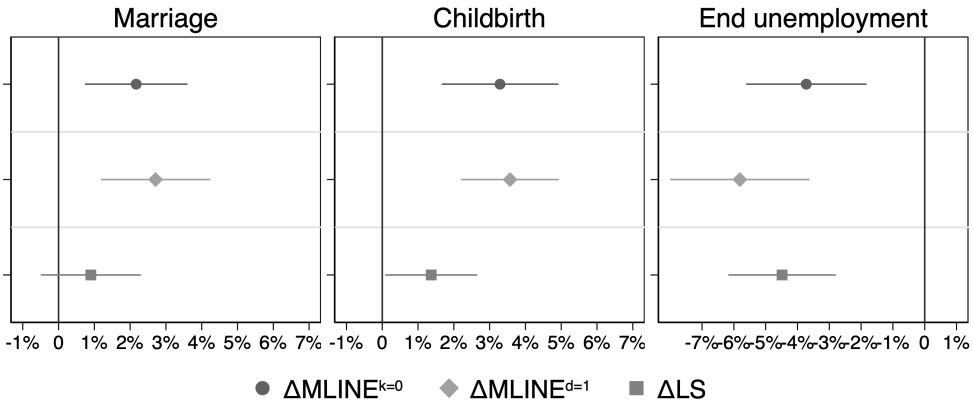


Figure A13: Robustness check: figure 5 with $\bar{m} = 0.24$

Reading note: Standardized OLS coefficients from the regression of an event on the lagged measure of well-being change. LS question: "How satisfied are you with your life, all things considered?" [0-10]. RLS question: "How satisfied were you a year ago with your life?" [0-10]. Source SOEP.

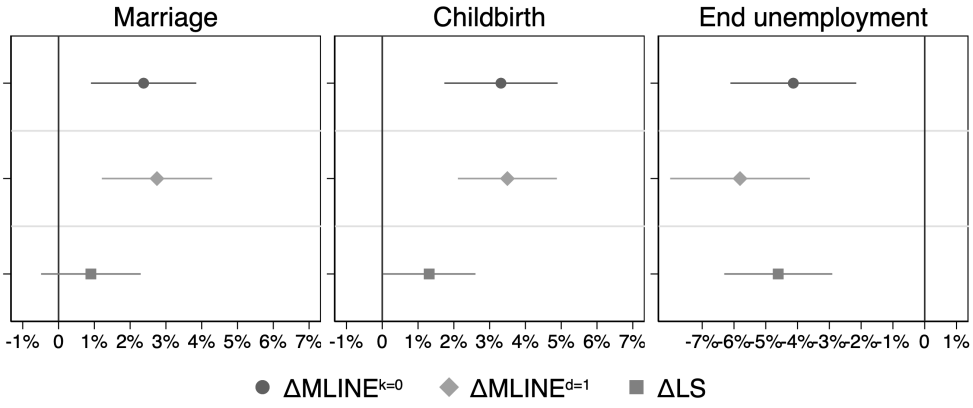


Figure A14: Robustness check: figure 5 with $\bar{m} = 0.6$

Reading note: Standardized OLS coefficients from the regression of an event on the lagged measure of well-being change. LS question: "How satisfied are you with your life, all things considered?" [0-10]. RLS question: "How satisfied were you a year ago with your life?" [0-10]. Source SOEP.

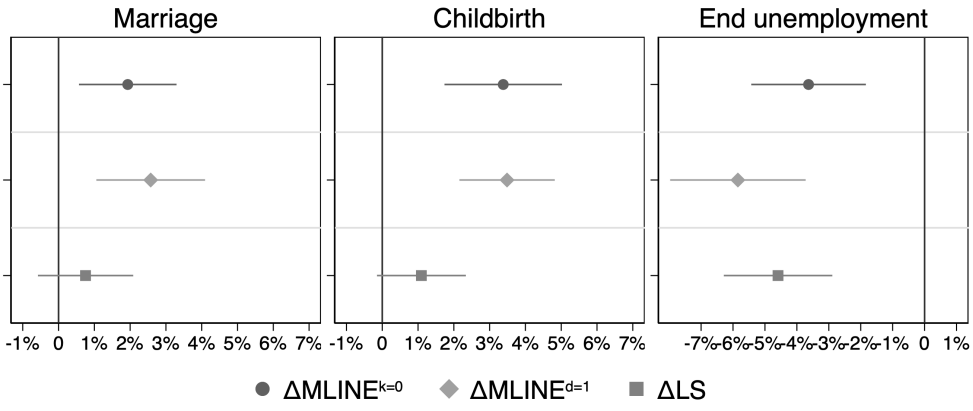


Figure A15: Robustness check: figure 5 without trimming

Reading note: Standardized OLS coefficients from the regression of an event on the lagged measure of well-being change. LS question: "How satisfied are you with your life, all things considered?" [0-10]. RLS question: "How satisfied were you a year ago with your life?" [0-10]. Source SOEP.

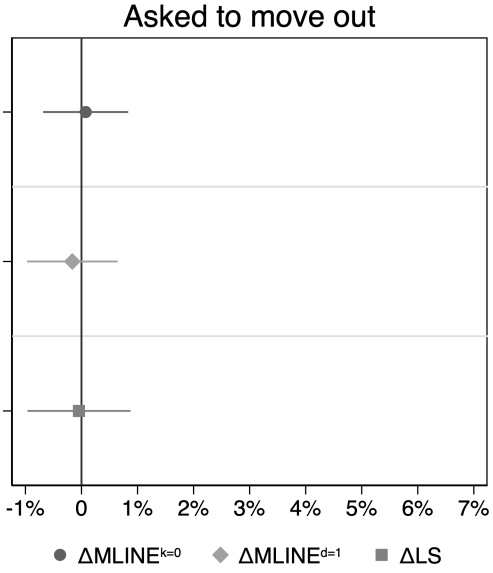


Figure A16: Placebo test: tenant's probability of being asked to move out next year

Reading note: Standardized OLS coefficients from the regression of an event on the lagged measure of well-being change. LS question: "How satisfied are you with your life, all things considered?" [0-10]. RLS question: "How satisfied were you a year ago with your life?" [0-10]. Source SOEP. We identify 53 events "being asked to move out" by combining information about the date of changing dwelling with the reason for it (*"Kündigung vom Vermieter"*).

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		
$\bar{m}^*$	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$^{k=0}$	6.59	7.43	8.43	9.75	10.75
			No stretch		
$\bar{m}^*$	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
M-LINE$^{d=1}$	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. $\bar{m}^*$ is a fixed parameter which was calibrated on the U.S. 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		
$\bar{m}^*$	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		
$\bar{m}^*$	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. $\bar{m}^*$ is a fixed parameter which was calibrated on the U.S. 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					
$\bar{m}^*$	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					
$\bar{m}^*$	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. $\bar{m}^*$ is a fixed parameter which was calibrated on the U.S. 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								
$\bar{m}^*$	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								
$\bar{m}^*$	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. $\bar{m}^*$ is a fixed parameter which was calibrated on the U.S. 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		
$\bar{m}^*$		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		
$\bar{m}^*$		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. $\bar{m}^*$ is a fixed parameter which was calibrated on the U.S. 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).

Top \ Bottom	Downward	No change	Upward	Total
Downward	10%	6%	5%	20%
No change	5%	30%	3%	38%
Upward	18%	15%	9%	42%
Total	32%	51%	18%	100%

Table A6: Joint distribution of perceived bottom and top rescaling

Reading note: Data from the U.S. online panel survey (N=1,627). For instance, 20% of respondents perceived that the top of the scale moved downward, while 18% of respondents perceived that the top of the scale moved upward. 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?".

Table A7: Robustness check for table 5: no trimming

	(1) ΔLS	(2) ΔLS	(3) ΔMLINE$^{k=0}$	(4) ΔMLINE$^{k=0}$	(5) ΔMLINE$^{d=1}$	(6) ΔMLINE$^{d=1}$
income change (log)	3.06	3.24	12.13***	12.13***	9.10***	9.10***
	(2.01)	(2.06)	(2.88)	(2.94)	(2.01)	(2.02)
childbirth	3.04	3.07	5.48**	4.76*	5.10***	4.81**
	(2.98)	(3.00)	(2.59)	(2.56)	(1.80)	(1.87)
new partner	4.06	6.17**	7.33**	8.19**	4.60	6.11*
	(2.95)	(3.03)	(3.32)	(3.31)	(3.23)	(3.43)
fired	-2.91	-1.75	0.94	0.50	-2.40	-2.53
	(2.95)	(3.21)	(5.19)	(5.18)	(3.21)	(3.29)
quit job	-0.72	-0.16	1.47	0.82	2.37	2.24
	(2.34)	(2.37)	(3.04)	(3.13)	(2.58)	(2.65)
start debt	-3.86**	-3.19**	-1.52	-2.39	-2.67*	-2.95**
	(1.57)	(1.57)	(1.55)	(1.58)	(1.46)	(1.45)
end debt	0.90	0.98	3.72***	3.08**	3.17***	2.88***
	(1.11)	(1.16)	(1.22)	(1.29)	(1.04)	(1.07)
new diagnosis/injury	-1.62	-0.08	-0.33	-0.20	-0.95	-0.92
	(1.15)	(1.34)	(1.12)	(1.24)	(0.94)	(1.08)
new job	2.92*	3.07*	6.41**	5.46**	4.30**	3.91**
	(1.77)	(1.77)	(2.88)	(2.72)	(1.69)	(1.68)
promoted	-0.22	-0.07	2.68	2.29	2.91**	2.94**
	(1.53)	(1.60)	(1.67)	(1.78)	(1.37)	(1.44)
bereavement/loss	-1.70	-1.32	0.72	-0.03	-0.49	-0.57
	(1.10)	(1.23)	(1.03)	(1.15)	(0.90)	(1.00)
Additional events		✓		✓		✓
N	1,006	1,006	995	995	1,006	1,006
R^2	0.027	0.056	0.116	0.142	0.105	0.121

Reading note: Linear regressions of 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. Source: U.K. online panel.

Table A8: Robustness check for table 5: $\bar{m} = 0$

	(1) ΔLS	(2) ΔLS	(3) ΔMLINE$^{k=0}$	(4) ΔMLINE$^{k=0}$	(5) ΔMLINE$^{d=1}$	(6) ΔMLINE$^{d=1}$
income change (log)	3.94**	4.17**	12.47***	12.33***	8.83***	8.61***
	(1.99)	(2.05)	(3.16)	(3.20)	(1.95)	(1.90)
childbirth	3.15	3.06	6.08**	4.95*	5.68***	4.72**
	(2.98)	(3.03)	(2.77)	(2.75)	(1.81)	(1.87)
new partner	2.59	4.81*	8.31**	8.73**	4.13	4.64
	(2.68)	(2.81)	(3.60)	(3.51)	(3.34)	(3.35)
fired	-3.02	-1.98	1.53	0.85	-2.05	-2.06
	(2.97)	(3.25)	(5.47)	(5.48)	(3.16)	(3.25)
quit job	-0.94	-0.49	2.45	1.90	2.73	2.56
	(2.41)	(2.43)	(3.27)	(3.33)	(2.33)	(2.39)
start debt	-3.28**	-2.59	-0.65	-1.61	-1.59	-2.07
	(1.58)	(1.58)	(1.62)	(1.64)	(1.32)	(1.35)
end debt	0.94	1.02	4.57***	3.74***	3.88***	3.13***
	(1.12)	(1.16)	(1.28)	(1.34)	(1.01)	(1.04)
new diagnosis/injury	-1.70	-0.12	-0.18	-0.40	-0.40	-0.50
	(1.15)	(1.34)	(1.11)	(1.25)	(0.94)	(1.07)
new job	3.40*	3.47**	6.74**	5.50*	3.79**	3.02*
	(1.76)	(1.75)	(3.18)	(2.99)	(1.65)	(1.58)
promoted	-0.50	-0.47	3.16*	2.41	3.47**	3.06**
	(1.52)	(1.59)	(1.76)	(1.88)	(1.36)	(1.44)
bereavement/loss	-1.87*	-1.61	1.40	0.27	0.24	-0.42
	(1.09)	(1.23)	(1.07)	(1.20)	(0.90)	(0.95)
Additional events		✓		✓		✓
N	988	988	988	988	988	988
R^2	0.028	0.060	0.132	0.160	0.123	0.148

Reading note: Linear regressions of 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. Source: U.K. online panel.

Table A9: Robustness check for table 5: $\bar{m} = 2.4$

	(1) ΔLS	(2) ΔLS	(3) ΔMLINE$^{k=0}$	(4) ΔMLINE$^{k=0}$	(5) ΔMLINE$^{d=1}$	(6) ΔMLINE$^{d=1}$
income change (log)	3.95**	4.17**	10.47***	10.41***	7.96***	7.82***
	(1.99)	(2.05)	(2.68)	(2.70)	(1.90)	(1.87)
childbirth	3.15	3.05	4.54*	3.91	4.42**	3.90**
	(2.98)	(3.03)	(2.46)	(2.47)	(1.74)	(1.83)
new partner	2.34	4.61*	7.00**	7.54**	3.90	4.52
	(2.61)	(2.77)	(3.16)	(3.13)	(3.27)	(3.32)
fired	-3.03	-1.99	0.68	0.33	-2.32	-2.18
	(2.97)	(3.25)	(4.91)	(4.93)	(3.19)	(3.26)
quit job	-0.93	-0.49	2.09	1.76	2.30	2.28
	(2.41)	(2.43)	(2.86)	(2.93)	(2.30)	(2.38)
start debt	-3.26**	-2.59	-1.49	-2.15	-2.27*	-2.56*
	(1.58)	(1.58)	(1.49)	(1.51)	(1.32)	(1.35)
end debt	0.95	1.01	3.32***	2.85**	2.85***	2.43**
	(1.12)	(1.16)	(1.17)	(1.23)	(1.01)	(1.04)
new diagnosis/injury	-1.66	-0.05	-1.47	-1.09	-1.75*	-1.39
	(1.15)	(1.33)	(1.03)	(1.14)	(0.94)	(1.07)
new job	3.41*	3.47**	4.98*	4.13*	2.78*	2.27
	(1.76)	(1.75)	(2.63)	(2.49)	(1.63)	(1.57)
promoted	-0.50	-0.47	2.49	2.20	2.80**	2.81*
	(1.52)	(1.59)	(1.59)	(1.69)	(1.37)	(1.44)
bereavement/loss	-1.89*	-1.62	-0.15	-0.64	-1.09	-1.18
	(1.09)	(1.22)	(0.98)	(1.09)	(0.89)	(0.93)
Additional events		✓		✓		✓
N	991	991	991	991	991	991
R^2	0.028	0.060	0.095	0.123	0.083	0.107

Reading note: Linear regressions of 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. Source: U.K. online panel.

Table A10: Robustness check for table 5: $\bar{m} = 6$

	(1) ΔLS	(2) ΔLS	(3) ΔMLINE$^{k=0}$	(4) ΔMLINE$^{k=0}$	(5) ΔMLINE$^{d=1}$	(6) ΔMLINE$^{d=1}$
income change (log)	4.03**	4.23**	7.90***	7.98***	6.64***	6.62***
	(1.98)	(2.05)	(2.22)	(2.25)	(1.87)	(1.85)
childbirth	3.15	3.05	2.38	2.45	2.54	2.69
	(2.98)	(3.02)	(2.15)	(2.24)	(1.71)	(1.87)
new partner	3.28	5.56**	5.61*	6.52**	3.56	4.52
	(2.51)	(2.65)	(2.87)	(2.89)	(3.43)	(3.50)
fired	-3.03	-2.00	-0.13	-0.18	-2.72	-2.36
	(2.97)	(3.25)	(4.34)	(4.36)	(3.33)	(3.36)
quit job	-0.97	-0.54	1.45	1.43	1.66	1.86
	(2.41)	(2.42)	(2.47)	(2.53)	(2.30)	(2.38)
start debt	-3.02*	-2.42	-2.65*	-2.85**	-3.33**	-3.30**
	(1.57)	(1.58)	(1.36)	(1.37)	(1.35)	(1.36)
end debt	0.83	0.87	1.64	1.72	1.34	1.42
	(1.11)	(1.15)	(1.06)	(1.12)	(1.02)	(1.06)
new diagnosis/injury	-1.72	-0.20	-3.43***	-2.25**	-3.89***	-2.81**
	(1.14)	(1.33)	(0.94)	(1.05)	(0.95)	(1.10)
new job	3.33*	3.37*	2.90	2.57	1.29	1.16
	(1.76)	(1.74)	(2.13)	(2.04)	(1.63)	(1.58)
promoted	-0.55	-0.47	1.46	1.75	1.81	2.42
	(1.52)	(1.59)	(1.45)	(1.52)	(1.42)	(1.47)
bereavement/loss	-1.84*	-1.56	-2.17**	-1.76*	-3.10***	-2.32**
	(1.09)	(1.22)	(0.90)	(0.98)	(0.91)	(0.94)
Additional events		✓		✓		✓
N	990	990	990	990	990	990
R^2	0.028	0.059	0.064	0.095	0.072	0.103

Reading note: Linear regressions of 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. Source: U.K. online panel.

Table A11: Robustness check for table 3: no trimming

	(1) ΔLS	(2) ΔMLINE$^{k=0}$	(3) ΔMLINE$^{d=1}$
income change (log)	0.26***	0.17***	0.17***
	(0.04)	(0.05)	(0.03)
childbirth	-0.17***	0.19***	0.06
	(0.06)	(0.07)	(0.04)
marriage	0.24**	0.61***	0.39***
	(0.10)	(0.11)	(0.07)
end unemployment	1.08***	1.91***	1.12***
	(0.16)	(0.24)	(0.13)
start unemployment	-0.81***	-0.48***	-0.73***
	(0.10)	(0.10)	(0.07)
new disability	-0.18	0.09	-0.14*
	(0.12)	(0.14)	(0.08)
bereavement	-0.60**	-0.04	-0.74**
	(0.29)	(0.53)	(0.31)
N	27,788	27,555	27,565
R^2	0.011	0.015	0.023

Reading note: Linear regressions of change individual changes in life satisfaction on life events. Column (1) uses nominal life satisfaction; column (2) uses the M-LINE under the no-shift scenario; column (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]. Source: SOEP core.

Table A12: Robustness check for table 3: $\bar{m} = 0$

	(1) ΔLS	(2) ΔMLINE$^{k=0}$	(3) ΔMLINE$^{d=1}$
income change (log)	0.25***	0.18***	0.20***
	(0.04)	(0.05)	(0.03)
childbirth	-0.15***	0.32***	0.16***
	(0.06)	(0.07)	(0.04)
marriage	0.21**	0.74***	0.49***
	(0.10)	(0.12)	(0.07)
end unemployment	0.92***	2.16***	1.14***
	(0.15)	(0.27)	(0.13)
start unemployment	-0.84***	-0.36***	-0.62***
	(0.09)	(0.11)	(0.07)
new disability	-0.14	0.24	-0.03
	(0.12)	(0.16)	(0.08)
bereavement	-0.60**	0.16	-0.68**
	(0.29)	(0.58)	(0.31)
N	27,485	27,485	27,485
R^2	0.011	0.017	0.024

Reading note: Linear regressions of change individual changes in life satisfaction on life events. Column (1) uses nominal life satisfaction; column (2) uses the M-LINE under the no-shift scenario; column (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]. Source: SOEP core.

Table A13: Robustness check for table 3: $\bar{m} = 0.24$

	(1) ΔLS	(2) ΔMLINE$^{k=0}$	(3) ΔMLINE$^{d=1}$
income change (log)	0.25***	0.12***	0.14***
	(0.04)	(0.04)	(0.03)
childbirth	-0.15***	0.08	-0.05
	(0.06)	(0.06)	(0.04)
marriage	0.21**	0.50***	0.30***
	(0.10)	(0.11)	(0.07)
end unemployment	0.92***	1.71***	0.94***
	(0.15)	(0.23)	(0.13)
start unemployment	-0.84***	-0.63***	-0.85***
	(0.09)	(0.10)	(0.07)
new disability	-0.14	-0.05	-0.26***
	(0.12)	(0.13)	(0.08)
bereavement	-0.60**	-0.22	-0.92***
	(0.29)	(0.48)	(0.31)
N	27,485	27,485	27,485
R^2	0.011	0.015	0.022

Reading note: Linear regressions of change individual changes in life satisfaction on life events. Column (1) uses nominal life satisfaction; column (2) uses the M-LINE under the no-shift scenario; column (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]. Source: SOEP core.

Table A14: Robustness check for table 3: $\bar{m} = 0.6$

	(1) ΔLS	(2) ΔMLINE$^{k=0}$	(3) ΔMLINE$^{d=1}$
income change (log)	0.24***	0.03	0.05*
	(0.04)	(0.04)	(0.03)
childbirth	-0.15**	-0.25***	-0.37***
	(0.06)	(0.06)	(0.04)
marriage	0.18*	0.17*	0.00
	(0.09)	(0.09)	(0.07)
end unemployment	0.87***	1.20***	0.66***
	(0.15)	(0.19)	(0.13)
start unemployment	-0.87***	-0.98***	-1.19***
	(0.09)	(0.09)	(0.08)
new disability	-0.15	-0.42***	-0.60***
	(0.12)	(0.11)	(0.08)
bereavement	-0.67**	-0.69*	-1.29***
	(0.29)	(0.40)	(0.31)
N	27,412	27,412	27,412
R^2	0.011	0.017	0.030

Reading note: Linear regressions of change individual changes in life satisfaction on life events. Column (1) uses nominal life satisfaction; column (2) uses the M-LINE under the no-shift scenario; column (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]. Source: SOEP core.

A3 Survey procedures

A3.1 U.S. 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 A15. 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 21 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 1,627 respondents.

	N	mean	sd	min	max
age	1,627	47.30	15.50	20	80
male	1,627	.49	.50	0	1

Table A15: Age and gender composition of the U.S. online sample

A3.2 U.K. online panel

The survey was programmed using Qualtrics and participants were recruited on Prolific (Palan and Schitter, 2018). Data collection took place in February-March 2023 (Wave 1) and February-March 2024 (Wave 2). Subjects were all U.K. 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 A16. 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, 16 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 990 respondents. Additional information about the data collection is available in Kaiser and Prati (2024).

In March 2025, we re-interviewed 771 participants, and, among them, 559 had already participated in both wave 1 and wave 2. We asked them to report information on their current and past income, and results are analyzed in section 2.

	N	mean	sd	min	max
age	990	42.95	13.27	18	82
male	990	0.49	.50	0	1

Table A16: Age and gender composition of the U.K. 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 $h_t \equiv \text{M-LINE}_t$:

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

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 $h_t^{d=1} \equiv \text{M-LINE}_t^{d=1}$:

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

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 yield bounds on 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:

$$d_t \cdot \frac{h_{t-1}}{h_{t-1} - k_t} = d_t \tag{A16}$$

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 A17: Conditions under which M-LINE$_t^{k=0} \leq h_t \leq$ M-LINE$_t^{d=1}$.

A4.3 Derivation of equations (14) and (15)

We report the derivation of the recursive expressions 14 and 15. For this, we start from the first two periods and then generalize.

First two periods

We begin with two expressions for latent satisfaction in period 1. From equation (6), we get the following expression for period 1:

$$h_1 = LS_1 \cdot d_1 + k_1 \tag{A23}$$

From the recursive update based on the M-LINE definition:

$$h_1 = LS_0 + (LS_1 - RLS_1 + \bar{m}) \cdot d_1 \tag{A24}$$

Where RLS_1 is short for $RLS_{1,0}$. Equating the two expressions for h_1 gives:

$$LS_1 \cdot d_1 + k_1 = LS_0 + (LS_1 - RLS_1 + \bar{m}) \cdot d_1 \tag{A25}$$

If $k_1 = 0$, then we can solve for d_1^*:

$$\begin{aligned} LS_1 \cdot d_1^* &= LS_0 + (LS_1 - RLS_1 + \bar{m}) \cdot d_1^* \\ LS_0 + \bar{m} \cdot d_1^* &= RLS_1 \cdot d_1^* \\ d_1^* &= \frac{LS_0}{RLS_1 + \bar{m}} \end{aligned} \tag{A26}$$

If $d_1 = 1$, then we can solve for k_1^*:

$$\begin{aligned} LS_1 + k_1^* &= LS_0 + (LS_1 - RLS_1 + \bar{m}) \\ k_1^* &= LS_0 - RLS_1 + \bar{m} \end{aligned} \tag{A27}$$

Any period t

If $k_t = 0$, then we can derive a general expression for d_t^* recursively. For example:

$$d_2^* = \frac{LS_0 \cdot LS_1}{(RLS_1 + \bar{m})(RLS_2 + \bar{m})}, \quad d_3^* = \frac{LS_0 \cdot LS_1 \cdot LS_2}{(RLS_1 + \bar{m})(RLS_2 + \bar{m})(RLS_3 + \bar{m})}$$

In general:

$$d_t^* = \frac{\prod_{n=0}^{t-1} LS_n}{\prod_{n=1}^{t}(RLS_{n,n-1} + \bar{m})} \tag{A28}$$

By harmonizing the time indexes in the product:

$$d_t^* = \prod_{n=1}^{t} \frac{LS_{n-1}}{RLS_{n,n-1} + \bar{m}} \tag{A29}$$

Finally, replacing $k_t = 0$ and d_t^* into the expression $h_t = LS_t \times d_t + k_t$, yields equation (14).

If $d_t = 1$, then, similarly, k_t^* satisfies:

$$k_2^* = LS_0 + LS_1 - RLS_1 - RLS_2 + 2\bar{m}, \quad k_3^* = LS_0 + LS_1 + LS_2 - RLS_1 - RLS_2 - RLS_3 + 3\bar{m}$$

In general:

$$k_t^* = \sum_{n=0}^{t-1} LS_n - \sum_{n=1}^{t} RLS_{n,n-1} + t\bar{m} \tag{A30}$$

By harmonizing the time indexes in the summation:

$$k_t^* = \sum_{n=1}^{t} (LS_{n-1} - RLS_{n,n-1}) + t\bar{m} \tag{A31}$$

Finally, replacing k_t^* and $d_t = 1$ into the expression $h_t = LS_t \times d_t + k_t$, yields equation (15).

A5 Making use of individual differences

Our paper develops the model in a "representative agent" framework, with our unit of observation being the average individual of a population (e.g. country). This section explores an extension of the model that uses the second moment of empirical distributions to identify the shift and stretch parameters at the same time. Although technically feasible, this new framework implies choosing somewhat arbitrary functional forms for the memory error, and adopting an array of assumptions which are hard to substantiate. These choices are not innocuous. While the pure "representative agent" framework that we develop in the paper is robust to various calibrations on the memory error, the framework developed in this section is not, since estimations are sensitive to the assumed structure of the variance of the memory error. We however discuss hereafter such a framework.[31]

A5.1 Variance of LS and RLS

Let the variables h_{it}, LS_{it}, and $RLS_{i,t,t-n}$ now all depend on individual i. Consider a single country, and the probability distribution over individuals in that country, so that the averages used in the analysis are means, e.g. $h_t = \mathbb{E}[h_{it}]$. Given a repeated cross-section, such expectations are identified (panel data is not necessary). If we ignore memory errors for the moment, we have:

$$LS_{it} = \frac{h_{it} - k_t}{d_t} \tag{A32}$$

$$RLS_{i,t,t-n} = \frac{h_{i,t-n} - k_t}{d_t} \tag{A33}$$

The above equations yield Equation (8) from the paper (written as Equation (A34) below):

$$\mathbb{E}[h_{it}] - \mathbb{E}[h_{i,t-1}] = (\mathbb{E}[LS_{it}] - \mathbb{E}[RLS_{i,t,t-1}]) \cdot d_t \tag{A34}$$

As in the main model, this identifies $\mathbb{E}[h_{it}]$ with the normalization $\mathbb{E}[h_{i0}] = \mathbb{E}[LS_{i0}]$ up to the parameter d_t. But notice that we can pin down d_t in this model directly by using the variance of LS:

$$sd(LS_{i,t-1}) = \frac{sd(h_{i,t-1})}{d_{t-1}}, \quad sd(RLS_{i,t,t-1}) = \frac{sd(h_{i,t-1})}{d_t} \tag{A35}$$

Therefore,

$$\frac{d_t}{d_{t-1}} = \frac{sd(LS_{i,t-1})}{sd(RLS_{i,t,t-1})}, \tag{A36}$$

which yields d_t for all t given $d_0 \equiv 1$. These standard deviations are knowable from the micro-data, and in some cases are even published in macro-level LS data.

[31] We would like to thank an anonymous referee for their inputs in developing this section.

A5.2 Reporting Function Heterogeneity

We develop our model with a single reporting function describing a population (e.g. country). However, it is worth noting that the single-agent model continues to hold even if the k_t are allowed to vary by individual k_{it}, insofar as d_t is maintained constant. Indeed, we can continue to identify d_t from:

$$\frac{d_t}{d_{t-1}} = \frac{sd(LS_{i,t-1})}{sd(RLS_{i,t,t-1})} \tag{A37}$$

since now:

$$sd(LS_{it}) = \frac{sd(h_{it} - k_{it})}{d_t}, \quad sd(RLS_{i,t,t-1}) = \frac{sd(h_{i,t-1} - k_{it})}{d_t} \tag{A38}$$

even if k_{it} and h_{it} are correlated. This model allows individuals to vary in their concept of h_{min} but not in their concept of $h_{max} - h_{min}$, since various additional restrictions would be necessary to identify $\mathbb{E}[d_{it}]$. Reporting function heterogeneity is not the focus of this paper and we do not develop it further. For a constructive investigation of the matter of heterogeneous reporting functions, see the recent contributions by Benjamin et al. (2023b) and Goff (2024).

A5.3 Variance of m

Let us now introduce the memory error, so that equation (A34) becomes:

$$\mathbb{E}[h_{it}] - \mathbb{E}[h_{i,t-1}] = (\mathbb{E}[LS_{it}] - \mathbb{E}[RLS_{i,t,t-1}] - \mathbb{E}[\bar{m_i}]) \cdot d_t \tag{A39}$$

If we assume that:

$$Cov(m_i, h_{it} - k_{it}) = 0, \tag{A40}$$

then:

$$\frac{d_t}{d_{t-1}} = \frac{sd(LS_{i,t-1})}{\sqrt{var(RLS_{i,t,t-1}) - var(m_i)}}. \tag{A41}$$

The variance of m is crucial for the estimation of the variance of h_{t-1} based on $RLS_{t,t-1}$. According to the law of recency, memory accuracy decreases over time, so that the variance of the memory error increases in a concave fashion. We assume a concave log-log-type function, with two parameters x and y:

$$var(m_t) = ln(x * ln(t + y)) \tag{A42}$$

where t represents the number of weeks elapsed between encoding and retrieval.

We know the variance of the error m based on a test-retest reliability measurement that we run during the same survey $(var(m_0)) = 0.64$ on the U.K. online sample and we know the variance of m after two weeks based on the U.S. online sample: $(var(m_{2weeks})) = 1.27$.[32] Based on this

[32]The data from the U.K. online sample was measured on a 1-7 scale and was therefore rescaled.

information, we can uniquely identify x and y from:

$$\begin{cases} \ln(x \cdot \ln(y)) = 0.64 \\ \ln(x \cdot \ln(2+y)) = 1.27 \end{cases} \tag{A43}$$

which yields ($x \approx 2.51$) and ($y \approx 2.13$).

Figure A17 depicts the predicted variance of the memory error from 0 to 52 weeks (there are 52.18 in one year). After one year, the predicted variance of m is $ln(2.51 * ln(52.18 + 2.13)) = 2.30$. After five years, it is $ln(2.51 * ln(52.18 * 5 + 2.13)) = 2.64$.

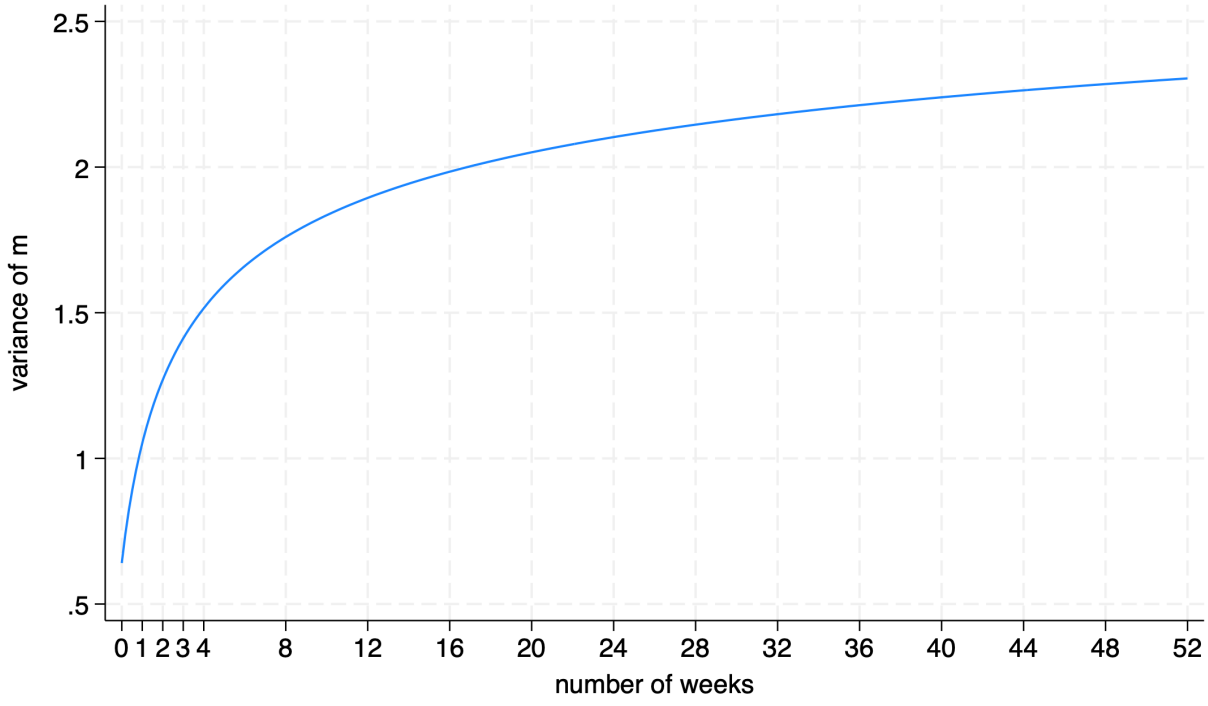


Figure A17: Variance of the memory error as a function of the retention interval

A5.4 From d_t/d_{t-1} to the M-LINE

Once the model is calibrated on a specific $var(m_i)$, we can rely on the empirical dispersions of $LS_{i,t-1}$ and $RLS_{i,t,t-1}$ to estimate d_t/d_{t-1}. And once d_t/d_{t-1} is identified, we can elicit the values of k_t and of latent satisfaction h_t. We can derive the difference between k_t and k_{t-1}:

$$\begin{aligned} RLS_{i,t,t-1} &= \frac{LS_{i,t-1} \cdot d_{t-1} + k_{t-1} - k_t}{d_t} - \bar{m} \\ d_t \cdot RLS_{i,t,t-1} &= LS_{i,t-1} \cdot d_{t-1} + k_{t-1} - k_t - d_t \cdot \bar{m} \\ k_t - k_{t-1} &= LS_{i,t-1} \cdot d_{t-1} - d_t \cdot RLS_{i,t,t-1} - d_t \cdot \bar{m} \end{aligned} \tag{A44}$$

And, finally, we can rely on k_t and d_t to calculate h_t for each period t, based on eq.(A34).

A5.5 Estimation of d with Recall Bias Heterogeneity USA (1959-2008)

Table A18 reports the summary statistics on the empirical dispersions in our historical dataset.

Period	LS_t	$sd(LS_t)$	RLS_t	$sd(RLS_t)$
1959-1963	6.58	2.24	5.89	2.68
1964-1968	6.81	2.05	5.92	2.52
1969-1973	6.49	2.12	5.62	2.55
1974-1978	6.56	2.25	5.55	2.65
1979-1983	6.48	2.08	5.72	2.50
1984-1988	6.48	1.95	5.77	2.39
1989-1993	6.44	1.93	5.96	2.26
1994-1998	7.10	1.80	5.70	2.50
1999-2003	6.80	1.82	6.08	2.22
2004-2008	7.05	1.81	6.37	2.13

Table A18: LS and RLS in the USA (1959-2008)

Following the previous section, we calibrate the model based on $\widehat{var(m_{5years})} = 2.64$. We now have all the information to calculate d_{t+1}/d_t:

$$\frac{d_t + 1}{d_t} = \frac{sd(LS_{i,t})}{\sqrt{var(RLS_{i,t+1,t}) - var(m_i)}}. \tag{A45}$$

Which yields the following values:

The table shows that the life satisfaction scale stretched by 240% over 50 years, averaging about 2.5% a year. Now that d_t/d_{t-1} is identified, we can proceed to identify d_t and k_t for all periods and, as a consequence, the M-LINE. Table A20 reports exactly this.

Overall, when we leverage the information about the dispersion of LS, RLS, and m, we can allow for stretching, shrinking and shifting at the same time. The M-LINE calculated this way suggests that latent satisfaction increased by about three times between 1959 and 2008, i.e., closer to the no-shift scenario than the no-stretch scenario.

However, it is worth reminding that these calculations are accurate only up to the accuracy of the calibration based on $\bar{m}$ and $var(m)$. The calculation of the ratio d_{t+1}/d_t is sensitive to the values of $\bar{m}$ and $var(m)$. For example, we need to rely on a specific calibration of the memory coefficient to claim that the life satisfaction scale expanded by 16% between 1963 and 1968. Not only the size of the stretch/shrink is affected by the variance of the memory error, but also the distinction between stretch and shrink is. Indeed, since d_{t+1}/d_t is very close to 1, a larger variance of the memory error could reverse the results whether $d_{t+1}/d_t > 1$ or $d_{t+1}/d_t < 1$, as the simulations reported in table A21 illustrates. This sensitivity makes this approach less robust empirically.

Period	d_t/d_{t-1}
1959–1963	1
1964–1968	1.16
1969–1973	1.05
1974–1978	1.01
1979–1983	1.18
1984–1988	1.19
1989–1993	1.24
1994–1998	1.02
1999–2003	1.19
2004–2008	1.31

Table A19: Stretch estimated by comparing LS_{t-1} and RLS_t (USA, 1959-2008)

Reading note: The table shows the estimated ratio d_t/d_{t-1}, i.e., the annual scale stretch. For instance, between the table says that, between the first and second period of the panel, the scale stretched by 16%. Estimations are calibrated using $\widehat{var(m)} = 2.64$. 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]. LS and RLS source: various datasets (cf table 1).

Period	d_t	k_t	M-LINE
1959–1963	1.00	0.00	6.58
1964–1968	1.16	-0.46	7.47
1969–1973	1.22	0.48	8.38
1974–1978	1.23	1.38	9.47
1979–1983	1.46	0.95	10.41
1984–1988	1.74	0.17	11.44
1989–1993	2.15	-1.65	12.21
1994–1998	2.19	-0.51	15.01
1999–2003	2.59	-1.06	16.58
2004–2008	3.40	-5.51	18.47

Table A20: Values of d_t, k_t, and M-LINE (USA, 1959-2009)

Reading note: The table shows the estimated values of d_t, k_t and of the $M - LINE$. In this model, all parameters are identified and we therefore do not need to distinguish between a no-stretch and a no-shift scenario. Estimations are calibrated using $\widehat{var(m)} = 2.64$. 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]. LS and RLS source: various datasets (cf table 1).

(1)	(2) d_t/d_{t-1}	(3) d_t/d_{t-1}	(4) d_t/d_{t-1}	(5) d_t/d_{t-1}
Calibration:	$0.5\widehat{var(m)}$	$0.75\widehat{var(m)}$	$1.25\widehat{var(m)}$	$1.5\widehat{var(m)}$
1959–1963	1	1	1	1
1964–1968	0.99	1.07	1.28	1.45
1969–1973	0.90	0.96	1.14	1.29
1974–1978	0.89	0.94	1.09	1.22
1979–1983	1.01	1.09	1.31	1.49
1984–1988	0.99	1.08	1.34	1.58
1989–1993	1.00	1.10	1.45	1.81
1994–1998	0.86	0.93	1.12	1.28
1999–2003	0.94	1.04	1.40	1.81
2004–2008	1.01	1.13	1.62	2.34

Table A21: Stretch estimated for different values of $var(m)$ (USA, 1959-2008)

Reading note: The table shows the estimated ratio d_t/d_{t-1}, i.e., the annual scale stretch, for different values of $\widehat{var(m)}$. Each column shows the annual scale stretch when the model is calibrated on a specific value, going from 50% of the original calibration (column (2)), to 50% larger than that (column (5)). The table illustrates that results change significantly based on the assumed value of the parameter, while this is not the case in the model developed in the main text. LS and RLS source: various datasets (cf table 1).

A6 Rescaling and socio-demographics

Rescaling is likely to be more impactful in conjunction with life events, hence its utility in unmasking fluctuations of latent satisfaction. A person who just had an income raise is likely to shift their standards upward, while someone who experienced a personal tragedy is likely to scale their benchmarks down. This is apparent in the retrospective happiness reported by paralyzed victims in Brickman et al. (1978) (see table 1 of their paper) and Kaiser (2022) comes to a similar conclusion when he finds that unemployment and widowhood particularly affect rescaling. But if some people rescale more than others - and in different directions – it is useful to know if they are systematically related with socio-demographic characteristics. In this paper, we refrain from heterogeneity analysis, but we hereby provide some tests that can help assess who tends to rescale more and in which direction.

We start with the U.S. online panel, where we collected information on the perceived movement of the top and bottom of the scale. We run two tests, using the only demographic variables which are available for all respondents, since they were used for stratifying the sample to achieve representativity, namely: age and sex. First, we run a logit regression where the dependent variable takes value 0 if the respondent perceived no change in the top or bottom of the scale over the past five years, and 1 otherwise. Second, we regress the answer to the perceived movement of the top (downward, none, upward) and bottom (downward, none, upward) in two separate regression models, using ordered probit. Table A22 shows the results. We find no difference between genders. Instead, older adults are relatively less likely to perceive a change in the scale, but more likely to see the top of their scale shrink than the rest of the population.

In a second exercise, we move on to the U.K. online panel, which contains both individual-level *estimations* of the rescaling parameters and a rich array of socio-demographic information. We can therefore run some tests to estimate if the shifting or the stretching parameters correlate with socio-demographic characteristics, in a similar fashion as perceived by the American respondents. In the first test, we regress the absolute value of the rescaling parameter (either k_i or d_i) on a set of socio-demographic variables which are standardly used in a life satisfaction regression, namely: age, sex, log(income), employment status, single/partnered, educational level, household size, number of children. The results, reported in table A23 shows that the size of rescaling is uncorrelated with all characteristics except for age. Older respondents tend to rescale less than younger ones, at least when it comes to shifting. The second test, reported in table A24, uses a logit estimation to detect any systematicity in the probability of upward rescaling (i.e., $k_i > 0$) compared to downward rescaling (columns (1) and (2)); and in the probability of stretching (i.e., $d_i > 1$) compared to shrinking. Results go in a similar direction as in the first test, that is, age comes up as reliably correlated with the type of rescaling. Specifically, the older people get the less likely they are to rescale upwards. Students and persons in non-standard professional situations (including: semi-retired, part-time students, carers) are also less likely to rescale upwards.

	(1)	(2)	(3)
	Any rescaling	Bottom	Top
age	-0.01***	0.00	-0.01***
	(0.00)	(0.00)	(0.00)
female	-0.00	0.03	0.03
	(0.11)	(0.06)	(0.06)
constant	1.51***		
	(0.19)		
cutoff 1		-0.44***	-1.19***
		(0.09)	(0.10)
cutoff 2		0.97***	-0.14
		(0.10)	(0.10)
N	1,619	1,619	1,619

Table A22: Perceived rescaling by socio-demographic (US, 2024)

Reading note: Regressions of perceived rescaling on age and sex. Column (1) estimates a logit regression where the dependent variable takes value 0 if the respondent perceived no change in the top or bottom of the scale over the past five years, and 1 otherwise; columns (2)-(3) estimate an ordered probit regression where the dependent variable takes value -1, 0 or 1 if the respondent perceive a movement downward, null or upward, respectively. Column (2) refers to the *bottom* of the scale, while column (3) to the *top* of the scale. Eight observations are missing because of non-response. Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	(1) $\|k\|$	(2) $\|k\|$	(3) $\|d\|$	(4) $\|d\|$
age	-0.09***	-0.07**	-0.00**	-0.00
	(0.02)	(0.03)	(0.00)	(0.00)
Male	-0.81	-0.80	-0.02	-0.01
	(0.62)	(0.67)	(0.02)	(0.02)
log(income)		-0.35		-0.01
		(0.48)		(0.02)
Not working		1.95		0.04
		(1.42)		(0.05)
Retired		-0.98		-0.04
		(1.42)		(0.05)
Self-employed		0.80		-0.00
		(1.21)		(0.04)
Unemployed		1.32		-0.01
		(2.27)		(0.08)
Student		0.16		-0.13*
		(1.97)		(0.07)
Other Professional		0.60		-0.06
		(1.38)		(0.05)
Partnered		-0.60		-0.03
		(0.94)		(0.03)
Medium education		-0.76		-0.01
		(0.83)		(0.03)
High education		-0.45		-0.02
		(0.83)		(0.03)
HH size = 2		-1.31		0.02
		(1.03)		(0.04)
HH size = 3		-1.31		-0.03
		(1.21)		(0.04)
HH size = 4		-0.73		0.05
		(1.57)		(0.06)
HH size = 5		-0.32		-0.07
		(2.45)		(0.09)
Nb of children = 1		0.57		-0.04
		(0.95)		(0.03)
Nb of children = 2		0.96		0.02
		(0.97)		(0.03)
Nb of children = 3		1.91		0.04
		(1.85)		(0.07)
Nb of children = 4		0.31		-0.18
		(4.91)		(0.17)
Constant	15.97***	19.18***	1.18***	1.31***
	(1.07)	(3.99)	(0.04)	(0.14)
N	983	929	983	929
R^2	0.016	0.029	0.006	0.023

Table A23: Size of rescaling by socio-demographic (U.K., 2023-24)

Reading note: Linear regressions of shifts and stretch in life satisfaction on socio-demographics. $N < 990$ because of missing observations, specifically: age (7 missing values), income (1), family composition (7) and education (46). Columns (1)-(2) use the absolute value of the shift parameter, i.e. $|k|$; columns (3)-(4) use the absolute value of the stretch parameter $|d|$. Baseline categories are: female, employed, low education, living alone, with no children. Low, medium and high education are defined, respectively, as: 0-2 years post-GCSE; 3-5 years post-GCSE, 6+ years post-GCSE. Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	(1)	(2)	(3)	(4)
	$\mathbf{1}_{\{k>0\}}$	$\mathbf{1}_{\{k>0\}}$	$\mathbf{1}_{\{d>1\}}$	$\mathbf{1}_{\{d>1\}}$
age	-0.01**	-0.02**	-0.02***	-0.02***
	(0.01)	(0.01)	(0.00)	(0.01)
male	-0.20	-0.21	-0.18	-0.21
	(0.15)	(0.16)	(0.13)	(0.14)
log(income)		0.09		0.06
		(0.11)		(0.10)
Not working		-0.20		-0.17
		(0.33)		(0.30)
Retired		0.24		-0.03
		(0.33)		(0.30)
Self-employed		0.02		-0.34
		(0.29)		(0.25)
Unemployed		-0.56		-0.26
		(0.49)		(0.48)
Student		-0.92**		-0.72*
		(0.44)		(0.42)
Other professional		-0.89***		-0.65**
		(0.30)		(0.29)
Partnered		0.05		0.30
		(0.22)		(0.20)
Medium education		-0.01		-0.18
		(0.20)		(0.18)
High education		-0.08		-0.27
		(0.20)		(0.18)
HH size = 2		0.40*		-0.03
		(0.24)		(0.22)
HH size = 3		0.57*		0.07
		(0.29)		(0.26)
HH size = 4		0.34		0.14
		(0.37)		(0.34)
HH size = 5		-0.08		0.11
		(0.54)		(0.53)
Nb of children = 1		-0.35		-0.07
		(0.22)		(0.20)
Nb of children = 2		-0.14		0.07
		(0.24)		(0.21)
Nb of children = 3		0.22		0.50
		(0.48)		(0.43)
Nb of children = 4		0.02		-1.60
		(1.18)		(1.18)
Constant	1.70***	1.12	1.28***	0.93
	(0.25)	(0.94)	(0.23)	(0.86)
N	983	929	983	929

Table A24: Probability of upward/downward rescaling by socio-demographic (U.K., 2023-24)

Reading note: Logit regressions of shifts and stretch in life satisfaction on socio-demographics. $N < 990$ because of missing observations, specifically: age (7 missing values), income (1), family composition (7) and education (46). Columns (1)-(2) use a binary variable = 0 if $k < 0$, and = 1 if $k > 0$; columns (3)-(4) use a binary variable = 0 if $d < 1$, and = 1 if $d > 1$. Baseline categories are: female, employed, low education, living alone, with no children. Low, medium and high education are defined, respectively, as: 0-2 years post-GCSE; 3-5 years post-GCSE, 6+ years post-GCSE. Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A7 Additional validation tests

A7.1 *Objective* reliability: test-retest reliability of RLS

If RLS is a reliable measure of past latent satisfaction, then the retrospective satisfaction of a particular period should not depend on the time it is measured, except for rescaling effects. We test this hypothesis using the U.S. online panel (N=1,627). We collect RLS scores at two time points that are too close to expect any rescaling (with a two-week interval: January 3, and January 17-20, 2024) and that refer to the same period. On both dates, respondents were asked to evaluate their life as of 5 years ago, using the Cantril ladder. On average, they gave the same response on the two dates ($RLS_{t,t-5y} = RLS_{t-2w,t-5y}$; diff=0.05; p-value = 0.175, N=1,627). This results lends some support to the reliability of retrospective satisfaction. Moving beyond simple means, figure A18 offers a scatterplot of the two RLS measures (one on each axis), showing that individual responses tend to cluster tightly around perfect agreement (i.e., the 45-degree line) rather than having dispersed answers that offset each other.

A7.2 *Subjective* reliability: perceived accuracy of RLS

One could wonder whether retrospective satisfaction does not merely capture 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 the second wave of our U.K. online panel, after asking participants to report their life satisfaction as of 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 self-reported $RLS_{t,t-1}$. Participants would then report a lower and upper bound for $RLS_{t,t-1}$. (The response was formulated as follows; "I am certain that my life satisfaction was actually: At the very least:...; At the very most:...".) We adopted the same procedure to estimate the amount of cognitive uncertainty about current life satisfaction LS_t, and compared the two.

Table A25 summarizes the results. The average individual in the sample is *certain* that their level of current life satisfaction is comprised between 56 and 73, and that their life satisfaction one year ago was between 54 and 70, on the 0-100 scale. The difference between the self-reported upper and lower bounds provides a measure of cognitive uncertainty, which is very similar in both cases (around 16-17 percentage points). This result shows that respondents feel fairly confident in how accurately they recall their past satisfaction - as confident as about their current life satisfaction.

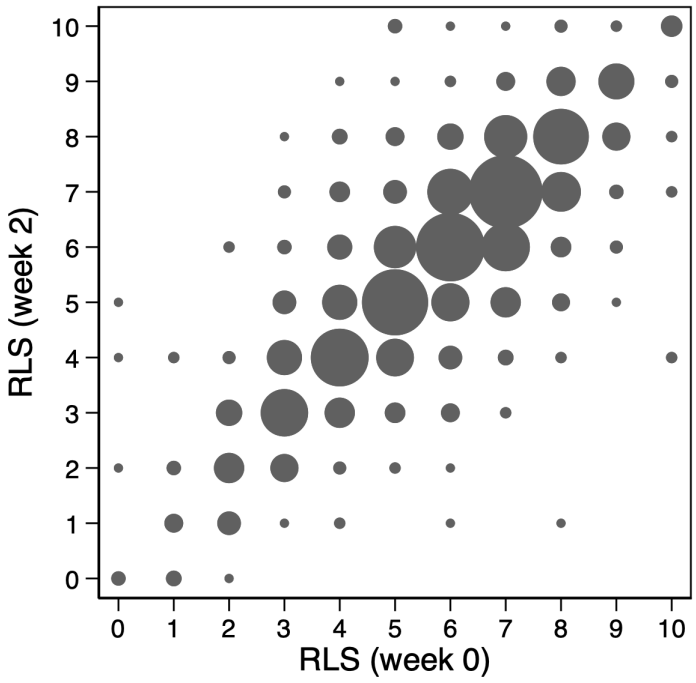


Figure A18: Joint distribution of RLS elicited two weeks apart.

Reading note: The scatterplot shown the distribution of RLS five years ago given on January 3rd 2024 (x-axis) and two weeks later (y-axis). The larger the bubble, the higher the density of responses. RLS question: "On which step would you say you stood five years ago?" [0-10]. Source: US online panel.

Table A25: Perceived accuracy of self-reports

	Lower bound	Upper bound	Cognitive uncertainty	N
LS_t	56.5	73.5	17.0	913
$RLS_{t,t-1}$	53.9	69.9	16.0	913

Reading note: The average respondent reports that they are *certain* that their current life satisfaction is between 56.5 and 73.5, and that their life satisfaction one year ago was between 53.9 and 69.9. The sample size is smaller than 990 because some responses failed the consistency check, that is, they reported a lower-bound that is higher than the upper-bound, of vice-versa. Source: U.K. online panel.