University of Oxford

Wellbeing Research Centre

Working Paper Series

2502

Scaling the Easterlin Paradox: Measuring Life Events on Stretching Happiness Measures

Charlie Harrison

July 2025

Scaling the Easterlin Paradox: Measuring Life Events on Stretching Happiness Measures

Charlie Harrison[1]

0) Abstract

Are people any happier than in the past? Despite rapid rises to GDP, average life satisfaction in various countries has not changed. Yet, richer people are happier than poorer people, and richer countries are happier than poorer countries. This is the Easterlin Paradox. Researchers typically explain the paradox by suggesting that these higher levels of wealth raise our expectations, preventing us from getting any happier (*the hedonic treadmill)*. The alternative explanation is that our wellbeing is increasing, but any improvements are being hidden by stricter reporting (*the rescaling hypothesis*). This second explanation has received little attention and researchers typically assume that no rescaling occurs. This paper proposes a new method to quantify rescaling effects. It analyses the short-run effects of five life major events on self-reported happiness, and how these effects have changed over time: a stretched reporting scale would result in falling absolute effect sizes. I test this idea using German panel data with ~650,000 observations (1991–2022). As the *rescaling hypothesis* predicts, the average effect of life events has fallen by ~35%. Due to scale expansion, self-reported happiness may be underreported by up to 50%. In other words, people may be living happier lives than in the past.

Words: roughly 7,400 (excluding Appendices, Bibliography)

My warm thanks to those who helped and offered comments on this article, especially Dr Alberto Prati, Dr Andrew Oswald, and Nathan Barnard.

[1] Oxford Wellbeing Research Centre, charlie.harrison.22@alumni.ucl.ac.uk. Code is available on request.

1) Introduction

Humans today live the wealthiest, longest, healthiest lives in history. Alongside increases in Gross Domestic Product, our societies are less violent, more educated, and more welcoming towards minorities than ever before (Pinker, 2018; Rosling, 2018). Yet, the mainstream view within wellbeing economics is that people are no happier for it. Average life satisfaction, assessed subjectively (e.g., from a 1 to 10 scale), is essentially constant in the long-run in various countries – from Germany and the UK, to China and India. Yet, in a given country, richer people are happier than poorer people, and richer countries, including those in Scandinavia, are happier than poorer countries. This apparent contradiction – income correlates with reported happiness at a point in time, but not in the long-run – is known as the *Easterlin Paradox*.

The paradox has two competing explanations. First, there is the *hedonic treadmill*. This suggests that our happiness has not increased because as we and our neighbours get richer, our expectations rise: comparison is the thief of joy. Second, there is the *rescaling hypothesis*. This suggests that our happiness has actually improved, but our reporting of it has become stricter (Prati & Senik, 2025). Despite the obviously important implications of the Easterlin Paradox – "*has the modern economy actually made anyone happier?"* – little work has been done to estimate rescaling effects.

In this paper, I propose a novel quantitative method to detect whether rescaling is occurring. *Rescaling* suggests that the happiness of a "10/10" life may be improving, whilst the quality of a "0/10" life remains the same. Just like measuring an object on an artificially stretched ruler, *rescaling* predicts that life events should have a *falling estimated effect over time*. This is exactly what we observe: the average effect of life events is converging towards zero. As an additional exercise, I reverse engineer happiness to adjust for rescaling, and find that due to stretching effects alone, latent happiness could be up to 50% higher than reported happiness. In short, the Easterlin Paradox may be due to underreporting.

This paper proceeds as follows. In Section 2, I review the relevant literature on the reporting function and rescaling. The two existing methodologies – using vignettes and memory – provide some evidence for rescaling, but both have limitations. Section 3 describes the methodology and within-person empirical approach. Section 4 describes the German panel data utilised. Section 5 presents my results. Section 6 discusses these results, including the implications for national happiness. Section 7 concludes with a summary, limitations, and room for further work.

2) Related Literature

2.1 The Easterlin Paradox: Two Competing Explanations

Easterlin (1974) observed that long-run national happiness was stagnant, despite steady rises to GDP. This finding constitutes a paradox for two reasons (Prati & Senik, 2025): first, it is inconsistent with many people's expectations; second, it seemingly contradicts other research which suggests that at a point in time richer people (and countries) are happier than poorer people (and countries). The Easterlin Paradox was initially based on US data, but has since been documented in Germany, UK, Australia, China, India, and Japan (Clark et al., 2018; Easterlin & O'Connor, 2022). While critics suggest that nominal happiness has increased (Stevenson & Wolfers, 2008), I do not address this debate here, and assume the paradox is true (or close to true).

Two theories attempt to explain the Easterlin Paradox. First, the *hedonic treadmill* suggests that underlying life satisfaction (LS) has not increased over time because LS is intrinsically relative. Prospect theory suggests happiness is shaped not by absolute income but by comparisons to reference points (Tversky & Kahneman, 1991). These benchmarks could increase due to hedonic adaptation, where expectations adjust upward as individuals and the people around them become wealthier. The *hedonic treadmill* can be integrated into a classical utility function by adding a term for one's reference income (Clark, Frijters et al., 2008).

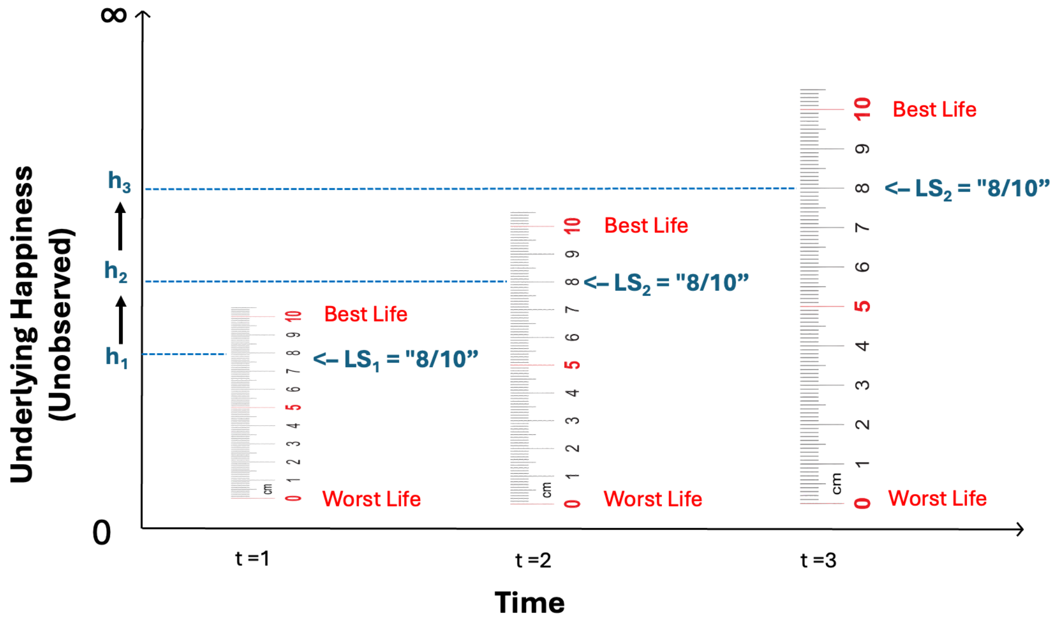
Alternatively, the *rescaling hypothesis* suggests that underlying happiness has increased over time but appears stagnant due to changes in measurement (Fabian, 2022a). This story centres around the *reporting function*: as proposed by Oswald (2008), this describes the relationship between the potentially boundless quantities of happiness we experience, and our numeric responses to LS survey questions with bounded integer scales. Researchers posit that people report their LS in roughly two stages (Fleurbaey & Blanchet, 2013). First, individuals decide which aspects of their life are relevant to determining their LS. Second, individuals compare their current life, along these particular dimensions, to other possible lives they might be living. As Kaiser (2022, p412) explains, one's LS response corresponds to the "quantile of one's life in a reference distribution".

We can therefore visualise the reporting function as a ruler placed in a space of unobserved happiness values (Figure 1).[2] Rescaling occurs if this reporting 'ruler' changes systematically over time. Figure 1 illustrates a *stretching* in the reporting function: the happiness of the "worst life" (h_w) remains relatively unchanged whilst the "best life" (h_b) improves. We could also conceive of *shifts* if h_w and h_b increase at the same rate.[3] Rescaling in either form, stretching or shifting, would lead reported LS to stagnate (e.g., at "8/10") despite increases in latent happiness.

[2] The reporting function, like a ruler, is likely to be linear: see *Section 6.3.* There are differences, however. The reporting function has fewer gradations than a ruler – there are typically only whole number responses to LS questions. The thresholds on the reporting function could either be 'soft' – i.e., rounding to the nearest whole number – or hard boundaries, as Kaiser (2022) describes.

[3] It is unclear whether we should expect more *stretching* or *shifting*. Research on quality of life suggests scale shifting but no stretching (Schwartz et al., 2018). Prati & Senik (2025) conducted interviews to explore changes in individuals' happiness scales over the past five years, and found that the most common response indicated stretching. Overall, it seems likely that if *rescaling* occurs, it involves both stretching and shifting.

Figure 1: The Rescaling Hypothesis’ Explanation for the Easterlin Paradox



Rescaling happens when we judge our lives against a shifting set of possible alternatives. What could cause our aspirations to evolve? There could be various individual-level reasons. One could be personal crises: Kahneman (1999) suggests that peoples’ apparent adaptation to spinal injuries are better explained by more modest benchmarks for assessing life satisfaction. Fabian (2022) argues that rescaling occurs when people intentionally *try* to compare their lives against worse alternatives: e.g., practicing gratitude for not being ill. Both these factors should not, however, lead to biased estimates of national happiness.

Another possible cause of rescaling is higher *economic* aspirations. Kahneman & Krueger (2006) suggest that higher utility may not increase life satisfaction because of an “aspiration treadmill”, biasing reporting. Prati & Senik (2025, pp. 9–10) give rescaling a distinctly material flavour: reversing the logic from Easterlin (1995), they suggest that economic aspirations could affect reporting behaviour, not underlying happiness.[4] If aspirations progressively increase as GDP increases, the quality of a “10/10” life could increase, leading to the stretching effects

[4] They also suggest non-material reasons for changing aspirations, including COVID-19, war, and having children.

illustrated in Figure 1.

To summarize: there are two explanations for the Easterlin Paradox. The *hedonic treadmill* suggests that happiness has stagnated because our material aspirations have risen; rescaling suggests underlying happiness has risen, but material aspirations have biased reporting downward. Both explanations share a similar causal structure and are hard to distinguish, because self-reported happiness reflects both underlying utility and how people report it (Ingelström & van der Deijl, 2021). The literature's reaction to rescaling effects has been highly polarised. Most researchers typically ignore the problem and assume no rescaling occurs (Plant, 2020). Others think measurement issues are much more serious, and conclude that self-reported happiness data cannot effectively guide policymaking (Fabian, 2022b; Whyte, 2013). Indeed, the work-horse methods in wellbeing science (e.g., fixed-effects models and ordered probits) assume common reporting scales, and therefore do not account for rescaling (Benjamin et al., 2023a; Kaiser, 2022). However, the extent of any bias from rescaling, and whether it can be corrected for, remain open questions, with emerging techniques seeking to measure rescaling effects (Plant, 2024).

2.2 Quantifying Rescaling Effects

Vignettes offer one approach: respondents rate the life satisfaction (LS) of a hypothetical individual described in a standardized scenario, and differences in ratings are interpreted as differences in scale use, allowing researchers to adjust LS scores accordingly. Studies using vignette adjustments typically find modest rescaling effects (Montgomery, 2022). However, this method relies on strong assumptions – that respondents interpret vignettes and their own lives using the same internal scale, and that all variation is due to reporting differences. People may interpret vignettes differently based on their own circumstances or temperament (Deaton, 2011). Vignettes are also costly and rarely included in large longitudinal surveys, making them ill-suited for evaluating long-term trends like the Easterlin Paradox.

An analogous approach is to use psychophysical calibration questions – for example, asking respondents to assess the brightness of light or the colour of shapes – to quantify differences in scale-use (Benjamin et al., 2023b). Again, however, psychophysical calibration questions are rarely included in historical datasets, so cannot be used to directly assess the Easterlin Paradox.

Memory-based methods provide an alternative by comparing individuals' recalled past happiness to contemporaneous ratings from that time. This "then-test" approach can detect rescaling if discrepancies arise between memory and historical data (Howard et al., 1979). These studies suggest rescaling may affect around 10% of respondents (Fabian, 2022a; Prati & Senik, 2022). Kaiser (2022) found, even with strong assumptions, that most coefficients retained the same size. Prati & Senik (2025) offer the only longitudinal correction for rescaling, estimating that LS in the US may be underreported by up to 140% in recent years. Yet, their results hinge on the assumption that memory biases are time-invariant. Recall-based methods face broader challenges, social desirability or motivated reasoning, when assessing historical happiness (Schwarz & Strack, 1999). There are, also, few long-run datasets with memory-based data.

Most of the literature on rescaling has focused on vignettes, psychophysical calibration, and memory-based methods, each with limitations. Instead, I propose a fourth, underutilized approach: using life events. Life events like bereavement likely correspond to real changes in peoples' underlying happiness: we have strong intuitions that they do, and they have noticeable effects on reported LS. If LS scores respond consistently to major life events over time, this suggests stable reporting functions – a more parsimonious explanation than assuming offsetting shifts in both reporting and true effects.

We can find similar reasoning elsewhere in the happiness economics literature. Prati and Senik (2022) and Helliwell et al. (2010) interpret consistency cross-time (via memories) and cross-country (via coefficients) as evidence against rescaling.[5] Similarly, Kahneman & Krueger

[5] Prati & Senik (2022) find that people accurately recall their previous happiness. They argue that consistent reporting standards and accurate memories is a more plausible explanation rather than faulty memories and rescaling effects cancelling each other out.

Helliwell et al. (2010) found that predictors like income, social trust, and governance quality have consistent effect sizes across countries. They conclude that the large international disparities in life satisfaction "are not due to different approaches to the meaning of a good life" (p11). Presumably consistent effects and consistent reporting functions may be more plausible than heterogeneities in

(2006) found similar effect sizes of life circumstances on net affect and life satisfaction, concluding there was "scant evidence" for rescaling, which they argue would only affect LS (not affect). To date, no study has directly used life events to calibrate life satisfaction over time. A partial exception is Andreoni et al. (2024), who find self-reported underreporting of negative events' emotional impact, but their design is cross-sectional and based on hypothetical judgments – not *actual experienced effects* of life events.

In sum, existing approaches either lack longitudinal reach or rely on strong assumptions. The field remains polarized in its treatment of rescaling – viewed as either trivial or paradigm-shifting– yet the evidence supports neither extreme. There is a clear opportunity to develop new, more robust methods, and life events offer a promising foundation for that effort.

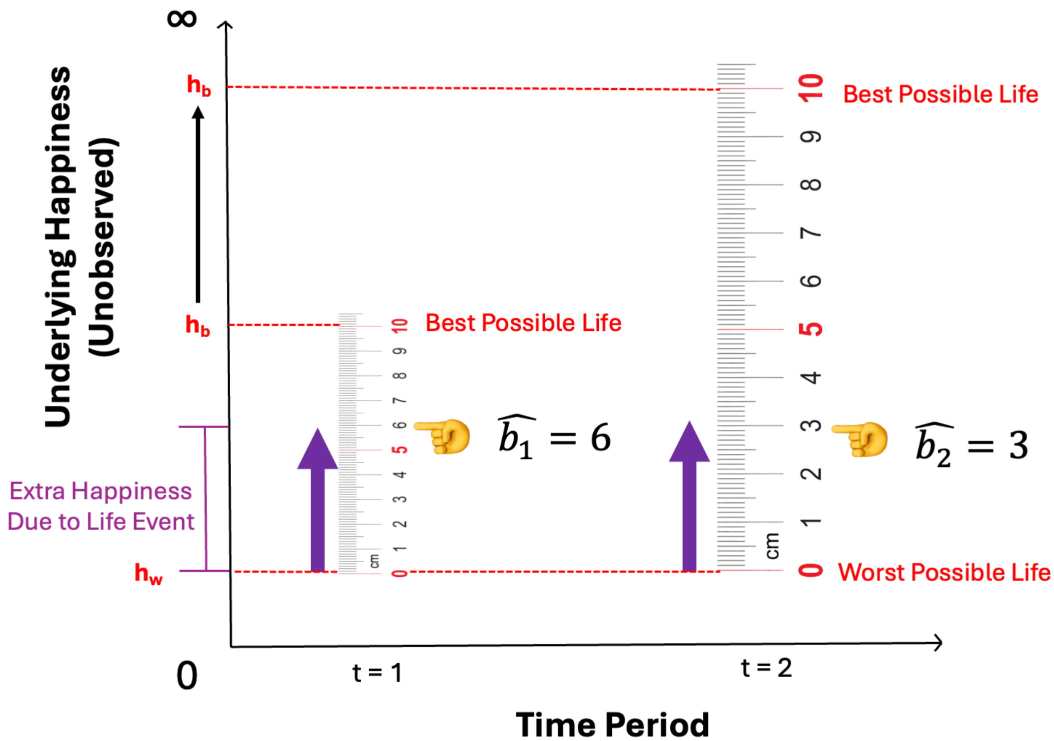
3) Methodology

3.1 Stretching, Attenuation, and Life Events

This paper examines the *rescaling hypothesis* using a novel approach: looking at whether the average effect of major life events on life satisfaction (LS) diminishes over time. Ultimately, I focus on five long-running life events: three relationship events, bereavement and unemployment. The basic intuition for this methodology is as follows. If rescaling is occurring, people compare their lives to an improving "best possible life", causing the gap between a "1/10" and a "10/10" life to grow and the reporting scale to stretch (*Section 2.2*). Suppose we measured the marginal amount of happiness associated with a given life event, for example unemployment. If the quantity of latent unhappiness associated with unemployment remains roughly constant, rescaling predicts that the estimated effects should attenuate over time. This is illustrated in Figure 2.

effects offsetting heterogeneities in reporting scales.

Figure 2: Visual Illustration of the Method for Identifying Rescaling



3.2 Baseline Model Specification

In my baseline specification, I estimate:

$$\Delta LS_{it} = \theta_0(Life\ Event_{it} * D_t) + e_{it}$$

The outcome variable, ΔLS_{it}, represents year-on-year changes in reported LS for a given individual ($LS_{it} - LS_{it-1}$), thus accounting for time-invariant confounders. While two-way fixed-effects (FE) remains the dominant approach, various studies on LS and life events employ first-differences (Ravallion & Lokshin, 2001; Graham et al., 2004). I opt for first-differences given the possibility that rescaling would introduce serial correlation in the idiosyncratic errors,[6] in which case the first-difference estimator would be more efficient (Wooldridge, 2013).

[6] Measurement error in △LS_it adds cumulatively to e_it (Wooldridge, 2010, pp. 71-72)

The independent variable – $Life\ Event_{it}$ – equals 1 if an individual experienced a major life event in the previous year. I look at short-run effects to maximise statistical power, as this is where the effect sizes on LS are typically largest (Clark et al., 2008).[7] Standard errors are clustered at the individual level to account for serial correlation and heteroskedasticity (Wooldridge, 2013, p511).

The central element in this paper is the interaction between the treatment and a set of dummy variables (D_t). This approach is widely used to analyze structural changes in treatment effects (Wooldridge, 2013, p453). By designating the first period (1991–1995) as the reference category, we can evaluate how treatment effects evolve over time. I use 5-year time bins to smooth out short-term fluctuations in effect sizes, such as those caused by the business cycle. Short-term noise would correspond to large relative changes, given that life events have small absolute effects (typically below 1 point on 0-10 LS scales). I conduct a sensitivity analysis using smaller (3-year) and larger (8, 10, and 15-year) bins.

3.3 Preferred Model Specification

The preferred model extends the baseline by introducing two key enhancements. First, it includes a set of time-varying covariates, X_{it}, drawn from Kaiser & Oswald (2022). These cover demographic and life-course characteristics such as age, family status, and job status. The full list is provided in Appendix 3. These covariates help address differential trends across individuals and improve the plausibility of the parallel trends assumption that underpins first-difference models. To further boost the plausibility of this assumption, both models restrict the control group for each life event to individuals who could plausibly experience that event, following Krämer et al. (2025).

Second, the preferred model includes interaction terms for two variables that may systematically shape the psychological impact of life events: age and event prevalence. Age is a potential

[7] By looking at the short-run effects of life events, I discount the possibility that life events cause rescaling over the long-run. Credit to Nathan Banard for this observation. As discussed in *Section 2.1*, I doubt that life events could bias national happiness estimates.

moderator of treatment effects, with older individuals typically exhibiting greater resilience and life satisfaction stability (Kubiszewski et al., 2020). Thus, we might expect age to moderate the effect of life events generally – although, for spousal bereavement, the sign may reverse, as older respondents become more vulnerable, beyond a certain age. To allow for non-linearities, banded age categories are used (18-24, 25-34, etc., 75+), with 45-54 as the reference category.

Social norms – which we proxy using the prevalence of each life event – may shape the emotional cost of those events. For example, a high prevailing unemployment rate may reduce stigma and expand social opportunities for the unemployed (Clark, 2003; Clark et al., 2010; Powdthavee, 2007). Similarly, Clark et al. (2018) found that, in some countries, the psychological cost of separation is greater when more people like you are partnered. Likewise, Wadsworth (2016) suggests the benefits of marriage may depend on its normative context. While bereavement research has focused primarily on other moderators – such as social resources or the length of the marriage in cases of spousal loss (Ghose et al., 2025) – it is plausible that social norms also moderate these effects. The mean event prevalence is used as the reference rate.

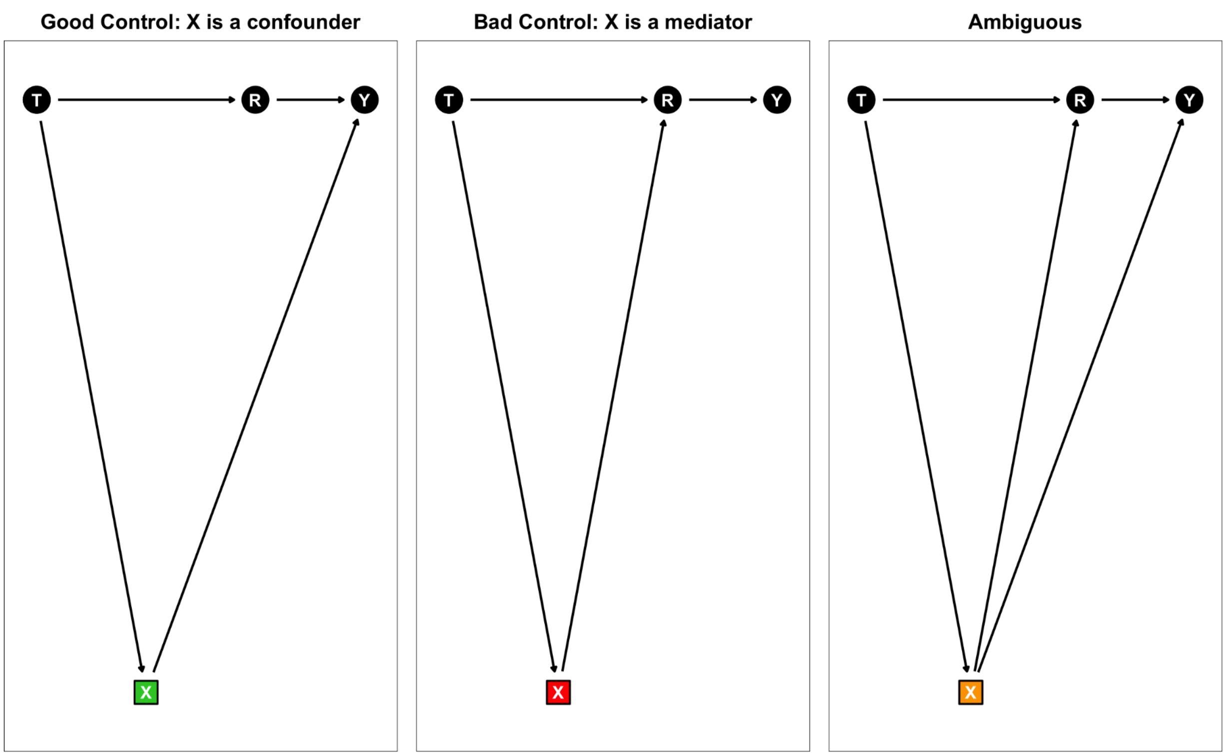
To capture social norms, interact each treatment variable with a five-year average of event prevalence, specific to each year, age group, and German region. For unemployment, this is the regional unemployment rate; for the three relationship events (new partnership, cohabitation, separation), it is the proportion of single individuals; and for spousal bereavement, it is the proportion of widowed individuals. These contextual prevalence rates are constructed from the SOEP separately for each of Germany's sixteen *Bundesländer*.

With both of these additional interactions, and socio-demographic controls, the resulting model is:

$$\triangle LS_{it} = \theta_0(Life\ Event_{it} * D_t) + \theta_1(Life\ Event_{it} * Age\ Category_{it}) + \theta_2(Life\ Event_{it} * Event\ Prevalence_{jt}) + X_{it} + e_{it}$$

This specification raises the possibility of over-control bias. As illustrated in Figure 3, while some covariates (X) may reduce confounding, if they lie on the path from time to treatment effects, other factors may influence treatment effects via rescaling itself. In the latter case, adjusting for X would block part of the causal mechanism and bias the estimate of interest (Pearl & Cinelli, 2022). Some variables may function as both confounders and mediators, in which case the direction of bias is unclear. Income-related variables – such as personal income, reference income, and hours worked – are potentially problematic in this respect. Given that they are closely tied to the existing literature on rescaling (*Section 2.1*), I exclude them from the control set. I assume that social norms around life events have little effect on the reporting function. Controlling for age may still introduce overcontrol bias (if income causes rescaling, since income rises with age). While our theory of rescaling remains unclear, the appropriate choice of controls is necessarily open to interpretation.

Figure 3: Good Controls; Bad Controls for Time (T), Unobserved Rescaling (R), and Treatment Effects (Y)



4) Data

4.1 Data Overview

This paper utilises the latest 32 waves from 1991 to 2022 of the German Socio-Economic Panel (SOEP), the longest running panel dataset with annual data on life satisfaction (Headey & Muffels, 2018). I exclude observations between 1984 and 1990 to ensure the dataset reflects post-unification Germany (Peichl & Ungerer, 2017).[8] The main analysis uses this unbalanced panel (with individuals entering and exiting at different times) in order to increase sample size (Kettlewell et al., 2020). Yearly cross-sectional, individual-level survey weights were used throughout.

4.2 Life Satisfaction and Selecting Life Events

The dependent variable is self-reported life satisfaction, which is assessed using the question: *"How satisfied are you with your life, all things considered?"* This question has been asked annually in SOEP since the start of the panel. Responses are recorded on an eleven-point scale (0–10), where 0 represents complete dissatisfaction and 10 represents complete satisfaction. This granularity is a key advantage of the German data: for example, the British Household Panel Survey uses a 7-point scale. Observations with missing data on life satisfaction are excluded.

The independent variables in this analysis are five major life events. These were selected from a list of 13 candidate events analysed by Krämer et al. (2025). Two criteria guided the selection, which is illustrated in Figure 4. First, data on each event had to be available from the start of the panel – the meta-analysis implemented later indexes effect sizes to those in the initial period (which must be consistent to ensure comparability). Second, each event needed to show a statistically significant effect, in a consistent direction, on life satisfaction in each 5-year period.

[8] Restricting the sample to West Germany across all years would have substantially reduced the sample size. An East/West dummy is included in the full controls specification; however, since treatment effects might vary by region, this would require interacting the treatment variable with this dummy. Interpreting the resulting three-way interactions would add considerable complexity to the analysis.

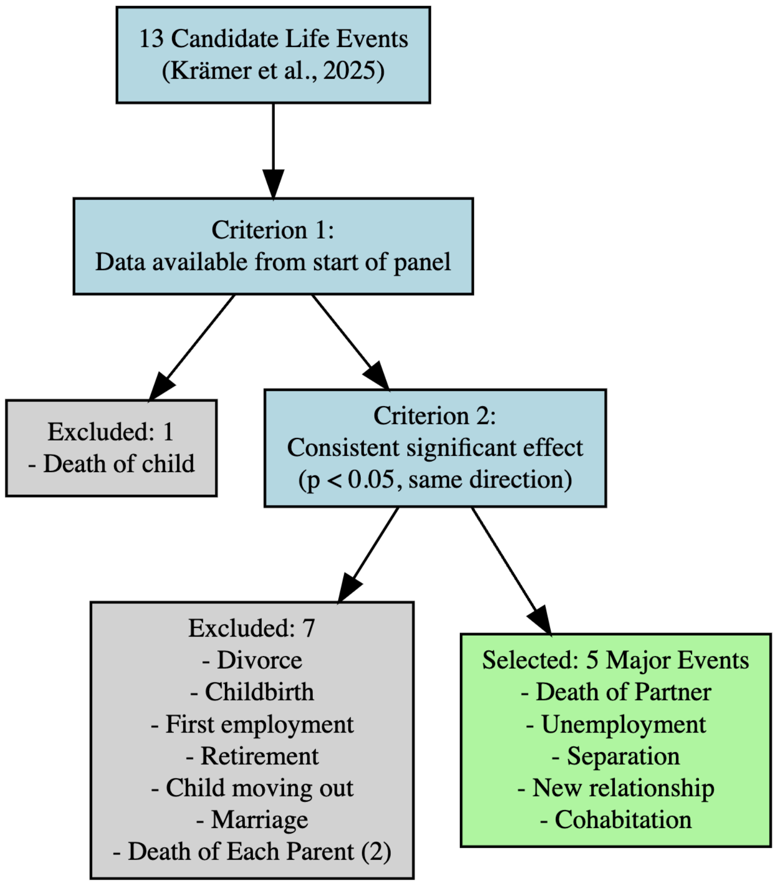
Specifically, the average marginal effect of the event had to be consistently significant at the 5% level in the same direction every 5-year period.[9] This criterion should increase precision without introducing bias.[10] Seven events were excluded on this basis. The final set includes five major life events: three with negative effects (death of partner, unemployment, and separation) and two with positive effects (starting a new relationship and cohabitation).

One possible concern is that the two positive events – forming a new partnership and beginning cohabitation – may be conceptually similar, as both involve romantic relationships. However, there are genuine conceptual differences. Cohabitation is a less frequent event (see Figure 5) and indicates a more committed partnership. In fact, cohabitation is around ten times more predictive of marriage in the following year than a new partnership (23% versus 2.4%). As a robustness check, I later group positive and negative events separately to account for potential collinearity.

[9] This does not assess the final shorter 3-year period (2020-2022), where several events (cohabitation, new partnership, separation) had non-significant effects. Within the exclusion stage, AMEs were computed using Stata's *margins* package, which evaluates the marginal effects in each 5-year period, averaging across each period's distribution of covariates (see *Stata Margins Reference Manual*, StataCorp, 2023: https://www.stata.com/manuals/cmmargins.pdf). In contrast, the meta-analyses described in *Section 5.3* uses interaction terms from the raw model outputs, which yielded slightly larger indexed standard errors.

[10] Consider a borderline event, with a small effect size. If the *rescaling hypothesis* is true, and the reporting scale stretches, the effect size might change from significant to insignificant; if the opposite prediction to stretching occurs (e.g., effect sizes grow), it might change from insignificant to significant. Excluding 'marginal events' should not stack the deck in favour of verifying rescaling.

Figure 4: Selection Criteria for Life Events

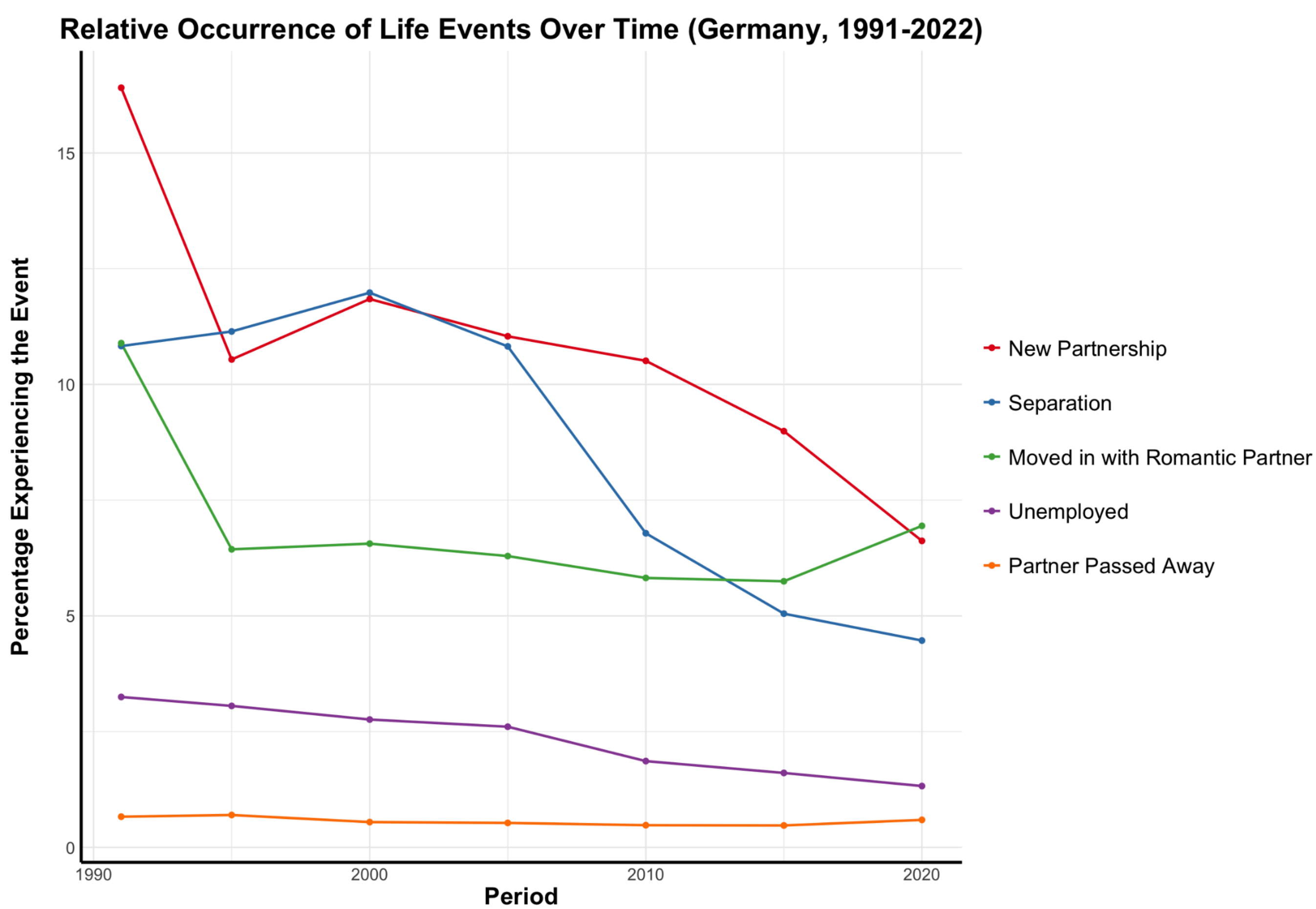


4.3 Defining Life Events

These events are coded as dummy variables (0 = event did not occur; 1 = event occurred) for each person-year observation. I use the same method as Krämer et al. (2025) to determine whether a life event has occurred, with full details in Appendix 3. Individuals for whom there is no life event data are re-coded as no event occurring, a reasonable approach since such events are relatively rare (Figure 5). As described above (*Section 3.3*), I restrict the control group to individuals who could plausibly experience the given life event. For example, for the death of one's partner, individuals are only included as controls if they are in a partnership. Further details are offered in Table 1.

I diverge from the data collection method in Krämer et al. (2025) on one major point: they distinguish between the first occurrence of an event and its subsequent occurrences – for example, treating the start of one's first new partnership separately from the second. Instead, my main analysis groups all occurrences together to maximize sample size, following Kettlewell et al. (2020) and Frijters et al. (2011). Additional sensitivity analysis focuses solely on first occurrences. Given this coding strategy and the inclusion conditions for the control group, I plot the relative counts for each event over time in Figure 6. The relative prevalence of each life event – aside from bereavement – seems to be falling over time, a point I will return to.

Figure 5:



5) Results

5.1 Individual Event Analysis

Of the five life events examined, there is evidence of attenuation for three (unemployment, cohabitation, and new partnership) and stability for the remaining two (separation, and bereavement). Amongst the attenuating events, whose time trends are illustrated in Figure 6, the clearest trends emerge for unemployment and cohabitation. For the former, the negative effect on life satisfaction moderates in more recent periods, with significant interactions emerging in 2015 in both the baseline and preferred models ($p \approx 0.01$). Chow tests, which assess the joint power of the period interactions, indicate temporal heterogeneity for both models ($p \approx 0.01$). Cohabitation shows a similar pattern. In the baseline model, the positive effect of moving in with a partner weakens over time, particularly in 2020-2022 ($p = .011$), with a marginally significant Chow test ($p = 0.050$). Attenuation is clearer in the preferred model, with significant negative interactions from 2000 onwards ($p < 0.01$) and a significant Chow test ($p = 0.004$). There are also signs of attenuation for new partnerships, particularly by 2005 ($p = 0.046$), but these effects are smaller and Chow tests remain non-significant ($p>0.10$).

By contrast, the effects of separation and spousal bereavement appear largely stable, as shown in Figure 7. Separation shows no evidence of temporal change, with no significant interactions or Chow tests under either model. The effects of spousal bereavement appear to *increase* under the baseline model – with one significant negative interaction in 2010 ($p = 0.049$). Magnification is the opposite trend to what *rescaling* predicts. However, this does not replicate in the preferred model, where interactions are weaker and non-significant ($p > 0.15$). The preferred model shows stable effects mainly because it controls for the increasing age of the sample. The negative effects of spousal bereavement are much higher within older age categories. Similar age interactions were observed for unemployment. For cohabitation and new partnerships, effects are more positive at younger ages, while separation has uniformly negative effects across all ages, with no significant moderation by age.

Allowing the effect of life events to vary according to their prevalence yields mixed results. Higher unemployment rates do not buffer the impact of unemployment ($p = 0.57$), and prevalence-based interactions for separation and bereavement are similarly weak. There is somewhat stronger, though still non-significant, evidence that the benefits of romantic relationships are greater where single rates are higher ($p \approx 0.1$ for new partnerships and cohabitation, respectively), as hypothesized.

In summary, attenuation over time is evident for unemployment, cohabitation, and possibly new partnerships. The effects of separation and bereavement shows no clear change over time, after accounting for age.

Figure 6:

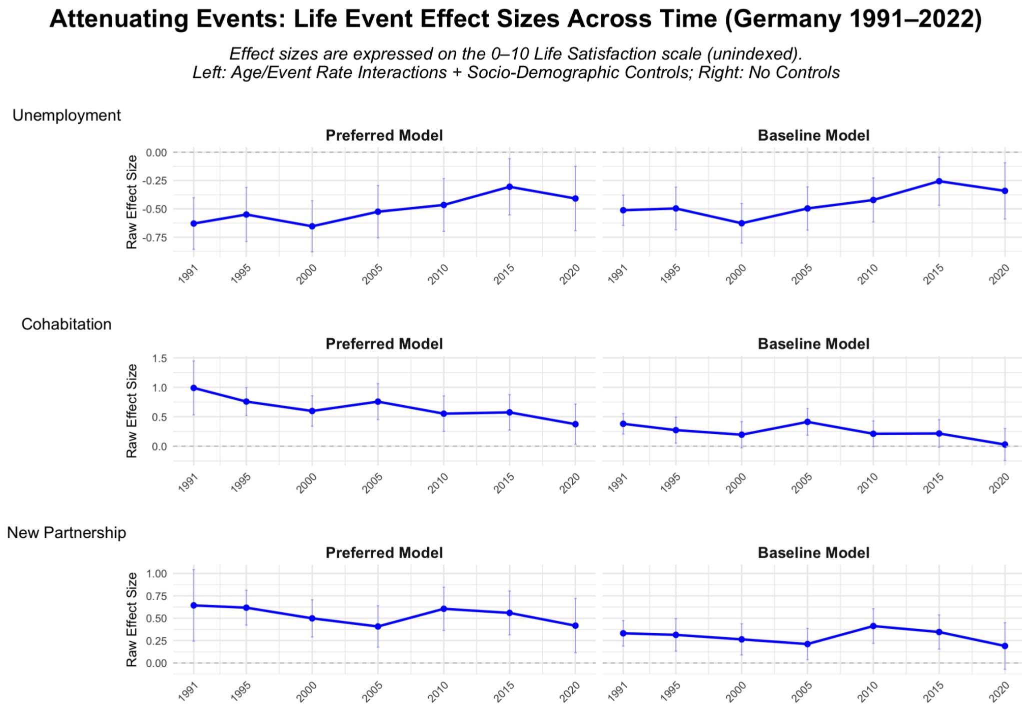
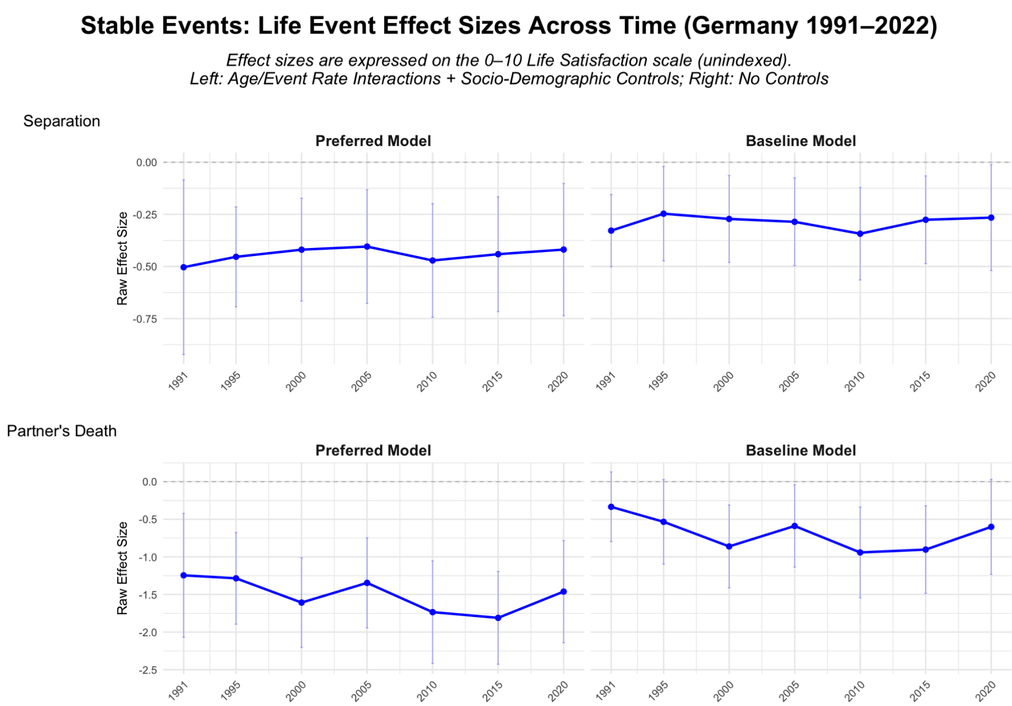


Figure 7:



5.2 Meta-Analysis of Life Event Effects

The initial results from the individual events are somewhat ambiguous: two events with attenuating effects, two events with stable effects, and one event with mildly magnified effects over time. What can we learn from this mixed picture?

To synthesize these findings across different events, I conducted a meta-analysis to determine whether the effect of *the average life event* fell over time. Like any meta-analysis, this process requires transforming effect sizes into comparable units. Currently, effects are expressed on an absolute 11-point life satisfaction scale. To standardize, I take the individual regression results from the preferred model, and index each event's effect size relative to its value in the first period. For example, the effect of unemployment fell from roughly -0.6 points to -0.3 using. This represents a 50% reduction in magnitude, so the initial indexed effect equals 1 and the 2015-2022 effect equals 0.5. A similar transformation was applied to the standard errors for the period-event interaction terms.[11]

The meta-analysis accounts for differences in precision across estimates. In line with standard practice (Lipsey & Wilson, 2001), I applied Hedges' weighting, which gives greater weight to more precisely estimated effects — calculated as the inverse of the squared standard error. Precision varied substantially by event. The average indexed standard error in the preferred model ranged from 13% of the initial effect for cohabitation to around 24% for separation: cohabitation was measured with the greatest precision and separation with the least. These differences in precision directly shaped the weights in the meta-analysis. Unemployment received the highest weight under the baseline specification (45%), followed by new partnerships (20%) and cohabitation (18%), with bereavement and separation contributing the least. Adding controls changed the precision pattern and reduced the weight assigned to unemployment to 20%, redistributing it primarily to cohabitation (35%), new partnerships (23%), and partner loss (12%).

[11] In this example: the interaction term for the final dummy was 0.3. Suppose this has a standard error of 0.1. This means that the standard error represents 16.67% of the effect of the initial effect (-0.6), so the indexed standard error would be 0.1667

I present the results of the meta-analysis for both the baseline and preferred models in Figure 8. Tables 1 and 2 present the same results, alongside alternative model specifications and data requirements (see also Appendix 2, Figures 13–14). Overall – in both the baseline and preferred models, the average effect of life events declines by approximately 35–40% from the first to the final period. This decrease is close to monotonic, particularly for the preferred model, consistent with gradual rescaling effects. Using the standardised standard errors, the meta-average falls significantly below 1 for both models in the last two periods (2015-2019, 2020-2022).

Other choices of period length yielded similar results (Appendix 2, Figure 11).[12] As an initial robustness check, I also computed a simple, unweighted average of the event effects across periods, which yielded a broadly similar pattern to the meta-average for the preferred specification (Appendix 2, Figure 12).[13] Other robustness checks are discussed in *Section 6.2*.

From the meta-analysis, we can conclude that the average effect of life events on life satisfaction is falling over time. Given this is what the *rescaling hypothesis* predicts, this constitutes reasonable evidence in favour of rescaling.

[12] Using shorter 3-year periods yielded noisier, but still attenuating average life event effects. The meta-average also declined using 8, 10, and 15-year periods, with significant falls in the final period (except for 15-year periods).

[13] The meta-average exhibits stronger attenuation than a simple-average under the baseline specification since the meta-analysis assigns less weight to spousal bereavement, which had a relatively higher indexed standard error (0.75 versus a typical 0.2-0.3).

Table 1:

Table 1: Indexed Life Events Effects in Germany, Relative to Base Period

Meta-Average across Events with 5-Years Periods
Base Period Effect = 1 (1991–1994). Inverse variance weights used for life events.

Period	(Baseline) Model 1	Model 2	Model 3	(Preferred) Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
1995–1999	0.904 (0.127)	0.883 (0.128)	0.910 (0.099)	0.870 (0.077)*	0.852 (0.076)*	0.834 (0.126)+	0.835 (0.077)*	0.917 (0.127)	0.929 (0.081)
2000–2004	0.994 (0.120)	0.964 (0.122)	0.960 (0.095)	0.825 (0.080)*	0.815 (0.079)**	0.850 (0.123)	0.806 (0.079)**	0.952 (0.123)	0.938 (0.081)
2005–2009	0.921 (0.126)	0.915 (0.127)	0.927 (0.099)	0.795 (0.088)**	0.790 (0.088)**	0.835 (0.126)+	0.807 (0.089)*	0.865 (0.123)	0.890 (0.091)
2010–2014	0.921 (0.130)	0.915 (0.132)	0.937 (0.103)	0.813 (0.090)*	0.802 (0.090)*	0.869 (0.128)	0.801 (0.093)*	0.938 (0.130)	0.890 (0.093)
2015–2019	0.737 (0.135)*	0.734 (0.137)*	0.831 (0.107)+	0.768 (0.091)**	0.760 (0.090)**	0.742 (0.132)*	0.740 (0.090)**	0.733 (0.127)*	0.822 (0.092)*
2020–2022	0.589 (0.163)**	0.576 (0.166)**	0.692 (0.127)**	0.646 (0.105)***	0.635 (0.104)***	0.557 (0.159)**	0.690 (0.108)**	0.594 (0.139)**	0.795 (0.107)*

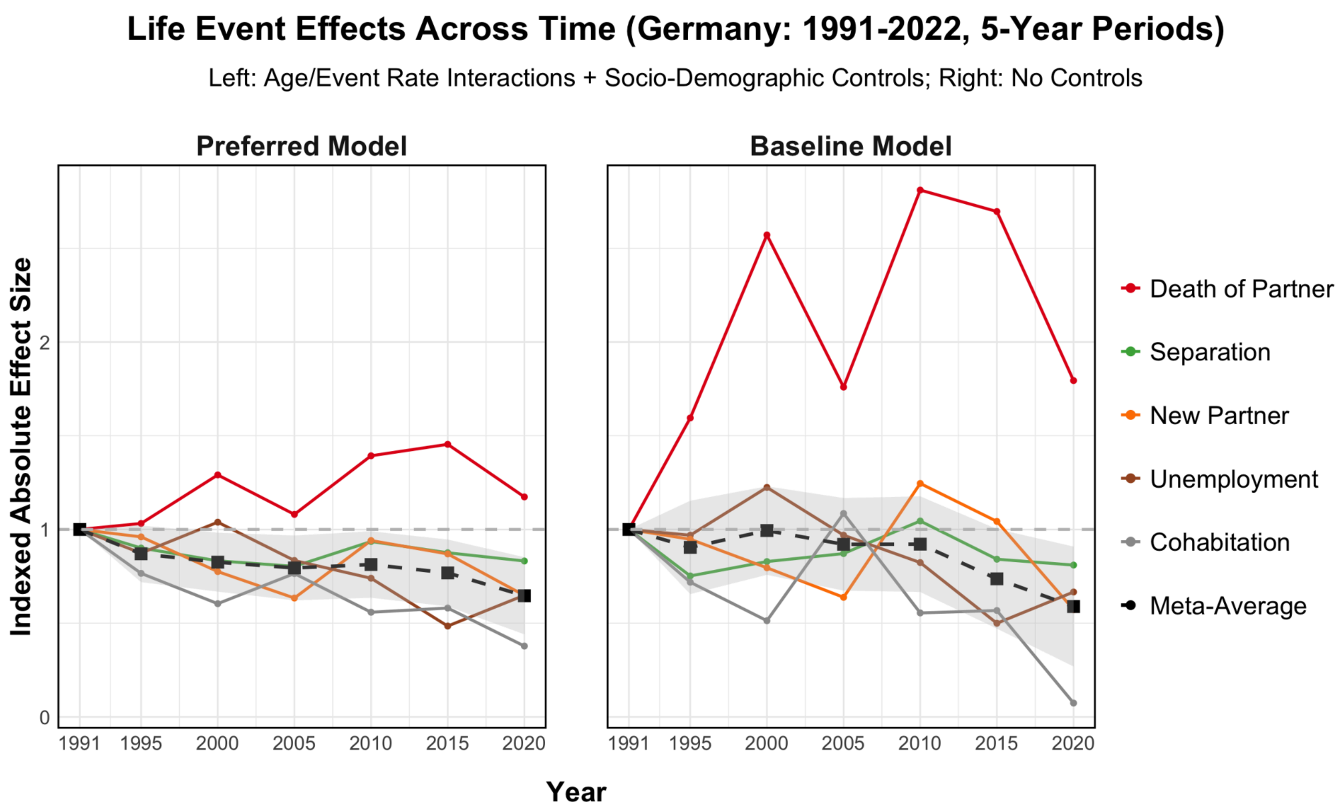
Note: One-sided p-values testing $H_0: \beta \geq 1$ vs $H_1: \beta < 1$. Significance: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 2:

Table 2: Model Specifications and Data Requirements

Feature	(Baseline) Model 1	Model 2	Model 3	(Preferred) Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Control Variables Included									
Socio-Demographic Controls?	✘	✔	✔	✔	✔	✔	✔	✔	✔
Interaction for Age?	✘	✘	✔	✔	✔	✔	✔	✔	✔
Interaction for Event Prevalence?	✘	✘	✘	✔	✔	✔	✔	✔	✔
Interaction for Income?	✘	✘	✘	✘	✔	✔	✘	✘	✘
Interaction for Reference Income?	✘	✘	✘	✘	✘	✔	✘	✘	✘
Data Requirements / Exclusions									
Non-Event Group Removed?	✔	✔	✔	✔	✔	✔	✔	✔	✔
1st Occurrence Only?	✘	✘	✘	✘	✘	✘	✔	✘	✘
Grouped Good/Bad Events?	✘	✘	✘	✘	✘	✘	✘	✔	✘
Exclude Already Treated?	✘	✘	✘	✘	✘	✘	✘	✘	✔

Figure 8:



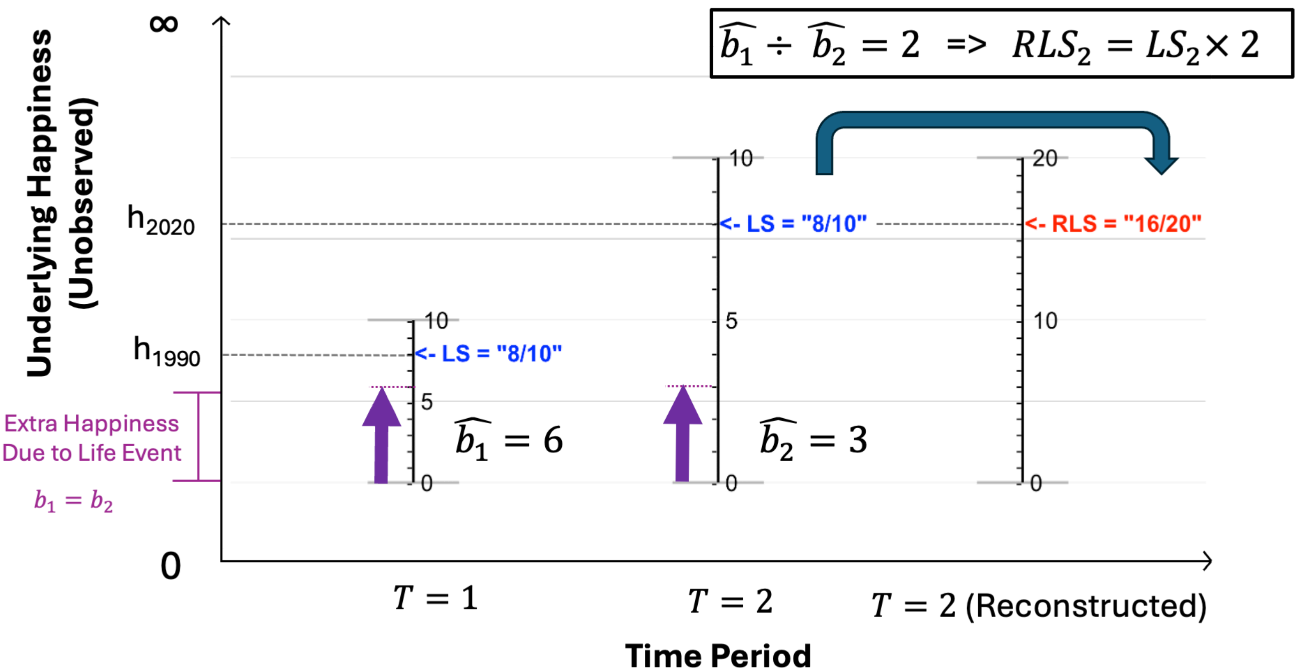
5.3 Adjusting Long-Run Life Satisfaction for Rescaling Effects

Next, under stronger assumptions, I reconstruct reported life satisfaction to adjust for rescaling effects. This involves "reverse engineering" LS, in the spirit of Prati & Senik (2025), who note a direct comparison to adjusting "current" dollars to "inflation-adjusted" dollars. To illustrate how this reconstruction process works, consider a toy example. Suppose the average effect of life events falls by half from $T = 1$ to $T = 2$. If the steps on the reporting scale are evenly spaced – like markings on a ruler – and, crucially, the true effect remains unchanged, then the LS scale must have expanded by a factor of 2 (***Proposition 1***). The intuition for this step comes from Plant (2020, p27).[14] Moreover, if, alongside this stretch, the bottom of the scale has stayed in the same place (i.e., no shifts), we can reconstruct the original scale by multiplying the reported LS in the second period by 2 (***Proposition 2***). I derive both propositions analytically in Appendix

[14] Plant (2020, p27) proposed this idea in the context of inter-cultural comparison.

1, with a novel proof which uses the reporting function model from Prati & Senik (2025). However, the process is easier to grasp visually, and this toy example is illustrated in Figure 9.

Figure 9: Toy Example of Using Life Events to Adjust for Rescaling

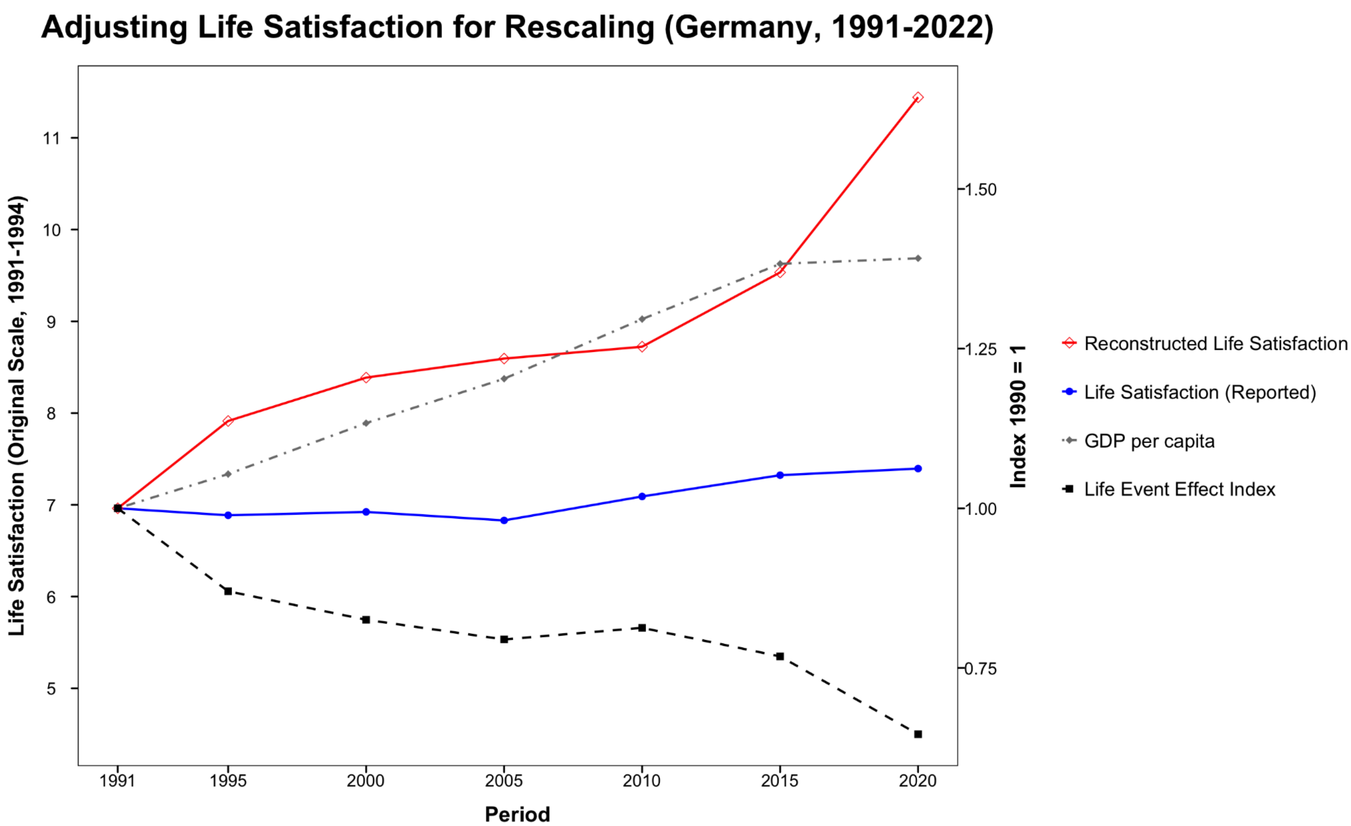


Using this methodology, I adjust average LS for rescaling using the same German data. In fact, Figure 10 represents a visual summary of the entire paper. Between 1991 and 2022, GDP per capita in Germany increased steadily, while average reported happiness remained relatively flat at around 7 – an illustration of the Easterlin Paradox. In red, I plot average national life satisfaction (LS) adjusted for rescaling, using the method just described. We found that the average effect of life events on reported LS declined by roughly 35% (using the preferred model) from the early 1990s to the early 2020s, implying substantial scale expansion. As a result, nominal LS could be underreported by 54% (a factor of 1/0.65) in recent years. In other words, average nominal happiness in the late 2010s (roughly 7/10) could actually refer to a score of roughly 11/10, when expressed on the 1990s scale.[15]

[15] More precisely – 7.395 ÷ 0.646 ≈ 11

This is how the *rescaling hypothesis* explains the Easterlin Paradox: once adjusted, reconstructed LS rises above reported LS from 2005 and rises faster than GDP from 2015 onward. In other words, people may be happier today than in the past, if we measure their happiness on a consistent scale. These results mirror Prati & Senik (2025), who found that, depending on assumptions, reconstructed LS was 80-140% higher than reported LS in the latest period.

Figure 10:



6) Discussion

This paper analyses the *rescaling hypothesis* by looking at whether the effects of life events. I found that, for several events, individual effects on life satisfaction have converged towards zero (1); the average effect of life events have fallen, implying rescaling is occurring (2); and that rescaling effects could lead to large underestimates of national happiness. Each stage involves increasing complexity and potential limitations, which I now turn towards.

6.1 Individual Life Events

There are at least two potential limitations to the individual event regressions: bias from heterogeneous treatment effects (HTE) and anticipation effects. First, within-person estimators like first differences (FD) can be biased if treatment effects vary over time, as they use already-treated units as controls (Goodman-Bacon, 2021). However, the proportion of never treated individuals is relatively high – from 72% (unemployment) to 89% (bereavement)[16] – and most life events (aside from unemployment) have minimal long-run effects (Clark et al., 2008), suggesting potential bias from heterogeneity is small (Baker et al., 2022). As a robustness check, I exclude already-treated units from the preferred model (Table 1, Model 9), and arrive at a comparable c20% fall in the meta-average.

A second potential limitation is anticipation effects, where LS changes in expectation of a life event. Event-study analyses typically interpret pre-treatment leads as anticipation rather than endogeneity (Clark et al., 2008). These effects usually share the same sign as short-run impacts, meaning within-estimators may underestimate the true treatment effects of life events (Malani & Reif, 2015). However, unless anticipation effects have grown, they would not lead to attenuation in effect sizes over time. The bias from heterogeneity and anticipation, while overlooked in wellbeing economics, is likely small.

6.2 Meta-Analysis

To assess how life event effects evolve over time on average, I conducted a meta-analysis. One potential limitation is that life events are not fully independent — correlations between them could bias the estimated effects. This presents two potential problems. First, co-occurence of events might bias the individual event regressions. Reassuringly, prior-event controls (e.g., Krämer et al., 2025) can address such dependencies, and this work suggests that effect sizes

[16] Proportion of never-treated units for other events: new partner (75%), cohabitation (80%), and separation (82%).

remain stable even in joint models, indicating that independence is a reasonable approximation. Second, the meta-analysis might duplicate information for events which are conceptually similar: e.g.,. cohabitation and new partnerships. As a robustness check, I coded binary variables for 'any good event', and 'any bad event', for each person-year, and then re-ran the meta-analysis. This yielded attenuation for positive events, negative events, and the meta-average (Table 1, Model 8), indicating that collinearity between events should not lead to bias.

Interpreting the meta-analysis is less straightforward. A decline in average effect sizes over time is consistent with the rescaling hypothesis, but two alternative explanations are possible: (a) statistical noise, and (b) falling true effects. I address each in turn.

6.2a) Rescaling or Statistical Noise?

Could the observed attenuation in life event effects be due to random variation? On one hand, the meta-average declined consistently across various model specifications and period lengths (Table 1). On the other hand, if rescaling is a general psychological process, one might expect a more uniform decline across all events. In particular, Figure 8 illustrates that the effect of bereavement (the red line) *increased* by around 20% from the first to the last period under the preferred model.

This concern is not fatal for *rescaling hypothesis*, and instead may reflect limited statistical power. The 95% confidence interval for bereavement in 2020-2022 ranges from a 72% *increase* to a 47% *decrease*.[17] The estimated effects of bereavement could have fallen if the sample size were larger. Further, the confidence intervals for both stable events (bereavement and separation) are generally larger. This difference in uncertainty is precisely what the meta-analysis aims to address. In short, the apparent stability (or slight magnification) in effect sizes for some events in some periods is not decisive evidence against rescaling. The 'pure chance' explanation appears insufficient to account for the clear downward trend in the meta-average.

6.2b) Rescaling or Falling True Effects?

[17] Precisely: 1.174 ± (1.96*0.278)

The second alternative is that the attenuation in observed effects is not due to rescaling – instead, it is because the true effects are falling. I suggest that this explanation cannot be ruled out definitively, but rescaling is more plausible.

First, there is no obvious theoretical reason for a decline in the psychological impact of various life events. The *hedonic treadmill* (Clark et al., 2008) for one addresses individual adaptation – not societal desensitisation. Whilst there is evidence that individuals become habituated to some life events after their first occurrence (Luhmann & Eid, 2009), attenuation persisted even after restricting the sample to the first occurrence of events (Table 1, Model 7). Second, the preferred model already controls for a wide range of structural and compositional shifts – including individual fixed effects (via first-differencing), the age of the respondent, and social norms (via the prevalence of the event). Significant attenuation also emerged prior to the COVID-19 pandemic.

While I cannot rule out all structural explanations, such as increased psychological support or the pace of modern life reducing sensitivity to life events, these ad hoc theories are difficult to test and unlikely to have been proposed *a priori*. This strengthens the case for rescaling as the more plausible explanation. Clearly, I cannot rule out all possible structural explanations. One could propose that wider access to psychological counselling or the increased pace of modern life makes the effect of life events smaller. These theories risk appearing ad hoc, and are often hard to test empirically. This strengthens the case for rescaling as the more plausible explanation.

One alternative explanation, which *is* testable, is that as people get richer, they become less sensitive to life events. For example, the negative effects of bereavement or unemployment could be mediated by having more savings. The attenuation in effect sizes persisted, or, if anything, *increased* after adding an income interaction to the preferred model (Table 1, Model 5). This is challenging for the *rescaling hypothesis*. If rescaling was due to rising living standards (*Section 2.2*), we would expect the controlling for income to nullify the attenuation, due to overcontrol bias (*Section 3.3*). Likewise, controlling for reference income, another proposed pathway, did not change the observed attenuation at all (Table 1, Model 6). This leads to a mixed conclusion. I suggest that rescaling is a more plausible interpretation for the results, than citing pure chance or residual structural effects. However, if rescaling is occurring, it is not following the typical model of rising material aspirations affecting reporting.

6.3 Adjusting LS for Rescaling

In the final step of this paper, using the results from the meta-analysis, I reconstructed LS to adjust for rescaling. This method relied on three key assumptions, which I will discuss in turn:

1. Reporting scale steps are evenly spaced: ‘linearity’
2. True life event effects remain constant: ‘variable equivalence’ (Plant, 2020)
3. The underlying happiness of a “0/10” life has remained the same: ‘no-shifts’

Linearity is the least problematic assumption, and is supported by a variety of empirical evidence. When asked to describe objective measures (e.g., height) using a subjective scale (e.g., 1-10), people do so linearly (Oswald, 2008). When assigning numerical values to normative labels (e.g., “very bad” to “very good”), they also space them evenly (Van Praag, 1991). And finally, the test-retest scores of LS have homoscedastic variance (Krueger & Schkade, 2008), which a non-linear reporting function would not predict.

However, even violations of linearity would not generally lead to the attenuation result. In order for non-linearity to lead to falling life event effect sizes, we would either require: (1) that reporting function has become more non-linear (which could be interpreted as a form of rescaling); or (2) that actual happiness has risen, in which case, rescaling must have happened given that reported happiness has stayed fairly constant in Germany (Figure 10).

The second assumption, *variance equivalence*, is unfortunately not empirically verifiable directly. Economists cannot directly measure peoples’ happiness. Whilst I suggest that rescaling is the best explanation for the observed attenuation, it is possible that the true effects are simultaneously falling due to some confounding factor. If this is the case, the reconstruction method could *overestimate* underlying happiness.

The *no-shift* assumption, while useful, is also problematic. Identifying *shifts* is simply not possible with this method: by analogy, while stretching a ruler would attenuate measurements, shifting it across the desk would not. It is possible that these shifts could be upwards – the happiness of a “0/10” life has increased – in which case, the reconstruction method would *underestimate* latent well-being. Alternatively, if downward shifting has occurred – a possibility the literature has not considered – the reconstruction would *overestimate* latent-wellbeing. In fact, if downward shifting is high enough (around 4 points), then overall national happiness could remain at a 7/10.

Overall, the reconstruction method offers a plausible *upper bound* on the bias results from a *stretching of the scale*. However, violations of the underlying assumptions – particularly the second and third – are possible, and could lead to bias in either direction. The overall level of national underlying happiness remains uncertain. I have therefore avoided reporting confidence intervals for reconstructed LS, and characterized the potential scale of misreporting as being on the order of 50%: perhaps greater than zero but inherently uncertain in magnitude.

7) Conclusion

I began asking the following question: do humans live happier lives than they did in the past? For various countries, long-run GDP has increased, while self-reported life satisfaction (LS) is stagnant. One neglected explanation for this paradox is that real improvements have been hidden by stricter reporting: this is the *rescaling hypothesis*. This paper empirically verified a prediction that rescaling makes: if the happiness gap between a “1/10” and a “10/10” expands, the estimated effect of life events measured along this scale should converge towards zero. The extent of the attenuation suggests that latent happiness may be underreported – potentially on the order of 50%. In other words, the Easterlin Paradox may be due to a stretching of the scales. In this concluding section, I suggest room for further work, as well as highlighting empirical and epistemic limitations.

As reviewed in Section 2, the academic literature on rescaling remains sparse. Only one study has previously attempted to adjust long-run LS data for rescaling effects. While this paper contributes new evidence, it cannot fully resolve the debate. Our conclusions must remain tentative – in Bayesian terms, we are working with highly uncertain priors and modest new evidence. One striking finding of this paper is that the attenuation of life event effects persisted even after controlling for material aspirations, a common mechanism proposed in the literature. Changes in peoples' aspirations may be occurring due to other mechanisms – e.g., social media comparisons or generational shifts in norms – which aren't captured in income statistics.

Several directions for future research emerge. The life event method could be applied to other long-running panel datasets, such as those from the UK or Australia. One could test a specific theory of rescaling by instrumenting a particular variable (e.g., reference income), and testing if it leads to more stable LS responses. Alternatively, researchers could test whether the R^2 of LS regressions (where happiness is the outcome) falls over time, reflecting a stretched reporting scale.[18]

This paper is not without limitations. Some are methodological and specific, whilst others are shared with wellbeing science more generally. First, the estimated effects of life events are small, making changes over time difficult to detect precisely. Second, the analysis focuses solely on German data, limiting generalizability. More fundamentally, my method for estimating rescaling effects relies on assumptions that are difficult to verify. First, it assumes that the true psychological impact of life events has remained stable over time. While I attempted to account for different socio-economic variables, residual structural factors remain possible. Second, the reconstruction method can only detect stretches in the scale, not shifts. The reconstructed LS series thus assumes that the happiness of a "0/10" life has not changed – an assumption that cannot be directly verified either.

Both challenges stem from a deeper issue: happiness is unobserved. This problem rears its head across the literature on the Easterlin Paradox. When richer people report higher happiness at a point in time, but long-run national happiness remains flat, is that because true happiness is stable, or because reporting has changed? When people remember being happier in the past, is that evidence of rescaling, or poor memory? In this study, when the effect of life events falls, is that due to rescaling — or is the true effects falling?

---

[18] Credit to Alberto Prati for this idea.

In each case, the competing explanations are observationally indistinguishable. Critics argue that this makes wellbeing research unscientific (Whyte, 2013). But this standard would surely exclude large swathes of conventional science. We cannot directly observe the early universe or the origin of life, yet we form robust scientific inferences by asking: which explanation best fits the data? This dissertation applies the same inferential logic. Given the significant attenuation in life event effects, I infer that the happiness scale has stretched over time, and that people may be happier today than they were three decades ago: the Easterlin Paradox may reflect a measurement illusion.

Appendices

Appendix 1: Proving the Reconstruction

A Model for the Reporting Function (Prati & Senik, 2025, p11-12).

Changes in the reporting function occur through two main mechanisms:

I) Scale Shifting (k_t): This occurs when both the upper and lower bounds of the reporting scale shift together. Conceptually, this means that both the best possible life (h_t^{max}) and worst possible life (h_t^{min}) increase over time. The shift parameter k_t can be thought of as the happiness level associated with the lowest point on the scale, such that: $k_t = h_t^{min}$

II) Scale Stretching (d_t): This occurs when the range of reported life satisfaction expands. If only the upper bound increases (i.e., h_t^{max} rises while h_t^{min} remains constant), this leads to upward stretching. The stretching parameter d_t is defined as the ratio of the true range of happiness to the fixed numerical range (0–10) of the LS scale:

$$d_t = \frac{h_t^{max} - h_t^{min}}{LS_{max} - LS_{min}} = \frac{h_t^{max} - h_t^{min}}{10} \quad (1)$$

This means that at any time t, an individual will declare a level of reported life satisfaction (0-10) of:

$$LS_t = \frac{h_t}{d_t} - \frac{k_t}{d_t} \quad (2)$$

This formulation assumes a linear reporting function (*Assumption I*). Next, we normalize happiness in the first period ($t_1 = 1$) as the reference point/numeraire:

$$h_1 \equiv LS_1$$

which implies from (2)

$$d_1 = 1;\ h_1 = 0$$

and that the reporting function in any subsequent period is represented by:

$$LS_t = \frac{h_t}{d_t} - \sum_{n=1}^{t}(k_n - k_{n-1}) \quad (3)$$

In simpler terms, this means: life satisfaction is reported by deflating underlying happiness by the stretching factor, and subtracting any cumulative shifts which have occurred.

Proving the relationship between Reconstructed LS, Latent LS, and Life Events

Let the effect of unemployment be the differences in means between the LS of the employed (LS_e) and unemployed (LS_u) – or, for any life event, the difference between the 'treated' and the 'untreated'. In the first period ($t_1 = 1$), this mean difference equals

$$\widehat{\beta}_{t=1} = LS_{e,1} - LS_{u,1} = h_{e,1} - h_{u,1}$$

since we are using the first period as the numéraire. Now, consider estimating this effect in a later period i:

$$\widehat{\beta}_{t=i} = LS_{e,i} - LS_{u,i}$$

From (3):

$$\widehat{\beta}_{t=i} = \left[\frac{h_{e,i}}{d_i} - \sum_{n=1}^{i}(k_i - k_{i-1})\right] - \left[\frac{h_{u,i}}{d_i} - \sum_{n=1}^{i}(k_i - k_{i-1})\right]$$

Assume k_i is identical across employed and unemployed, then:

$$\widehat{\beta}_{t=i} = \frac{h_{e,i} - h_{u,i}}{d_i}$$

To isolate d_i, we divide the treatment effect in period i by the treatment in the first period:

$$\frac{\widehat{\beta}_{t=i}}{\widehat{\beta}_{t=1}} = \frac{h_{e,i} - h_{u,i}}{d_i} \times \frac{1}{h_{e,1} - h_{u,1}}$$

Assuming that the true treatment effect remains constant (Assumption 2):

$$d_i = \frac{\widehat{\beta}_{t=1}}{\widehat{\beta}_{t=i}}$$

=> Proposition 1: the stretching parameter in period i equals the ratio of life event coefficients, between $t = 1$ and $t = i$

Rearranging (1).

$$h_i = (LS_i \times d_i) + [d_i \times \sum_{n=1}^{i}(k_i - k_{i-1})] \qquad (4)$$

If there are no shifts in reporting scale (Assumption 3), sum of first differences of k_n equal 0, and

$$h_i \approx (LS_i \times \frac{\widehat{\beta}_{t=1}}{\widehat{\beta}_{t=i}}) \qquad (5)$$

=> Proposition 2: Life Satisfaction in period i, expressed in the initial reporting scale, equals reported LS multiplied by the ratio of coefficients.

For example, suppose $\widehat{\beta}_{t=1} = 6$, $\widehat{\beta}_{t=i} = 3$

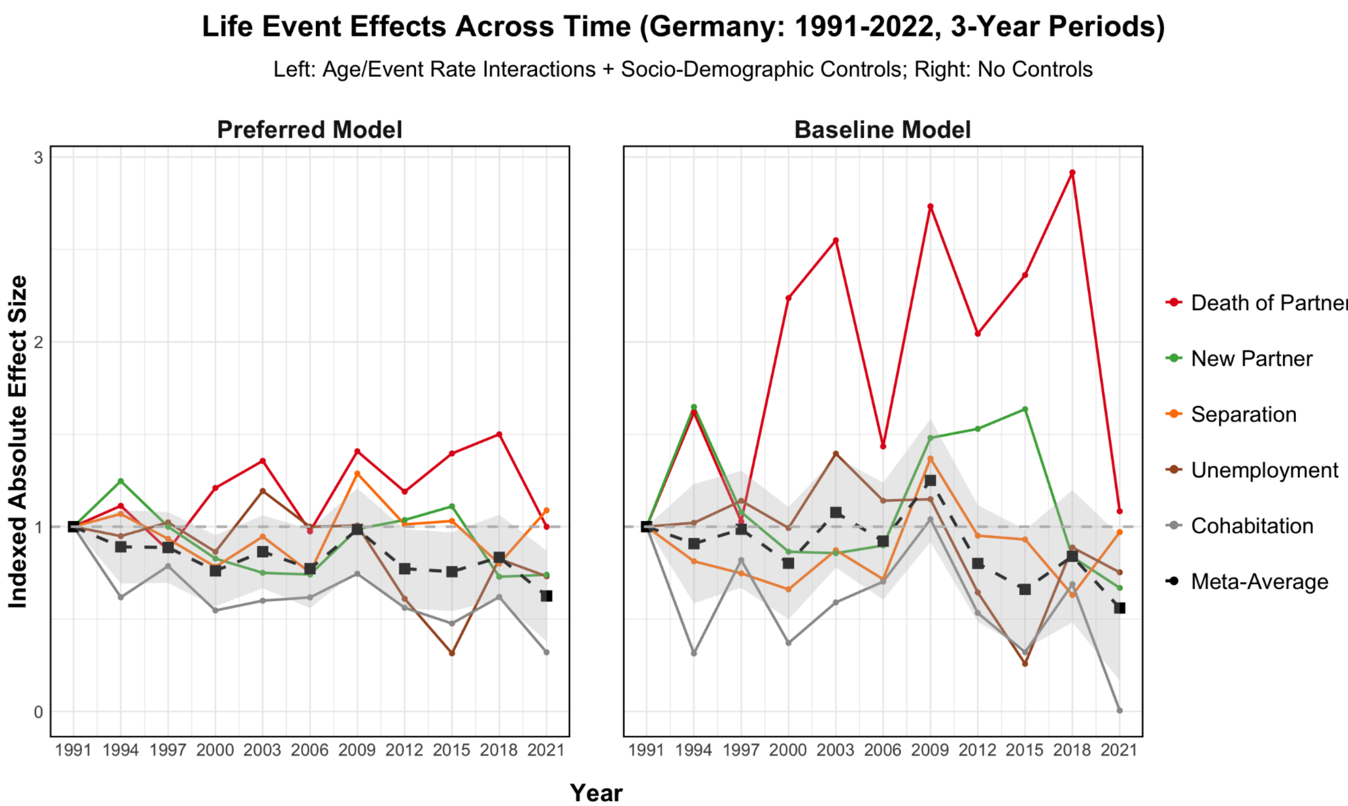
$$h_i \approx (LS_i \times 2)$$

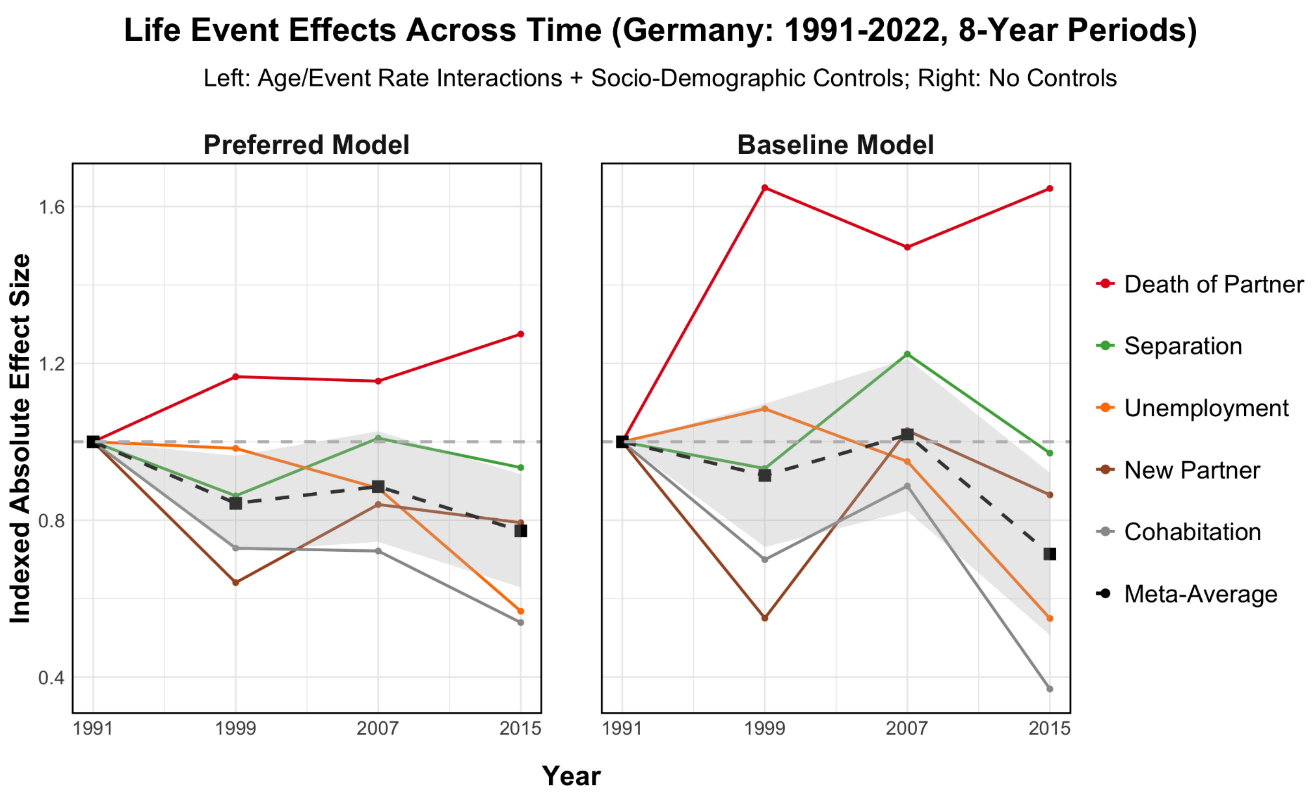
Thus, if the coefficient on life events halves, a given level of reported Life Satisfaction refers to double the underlying happiness.

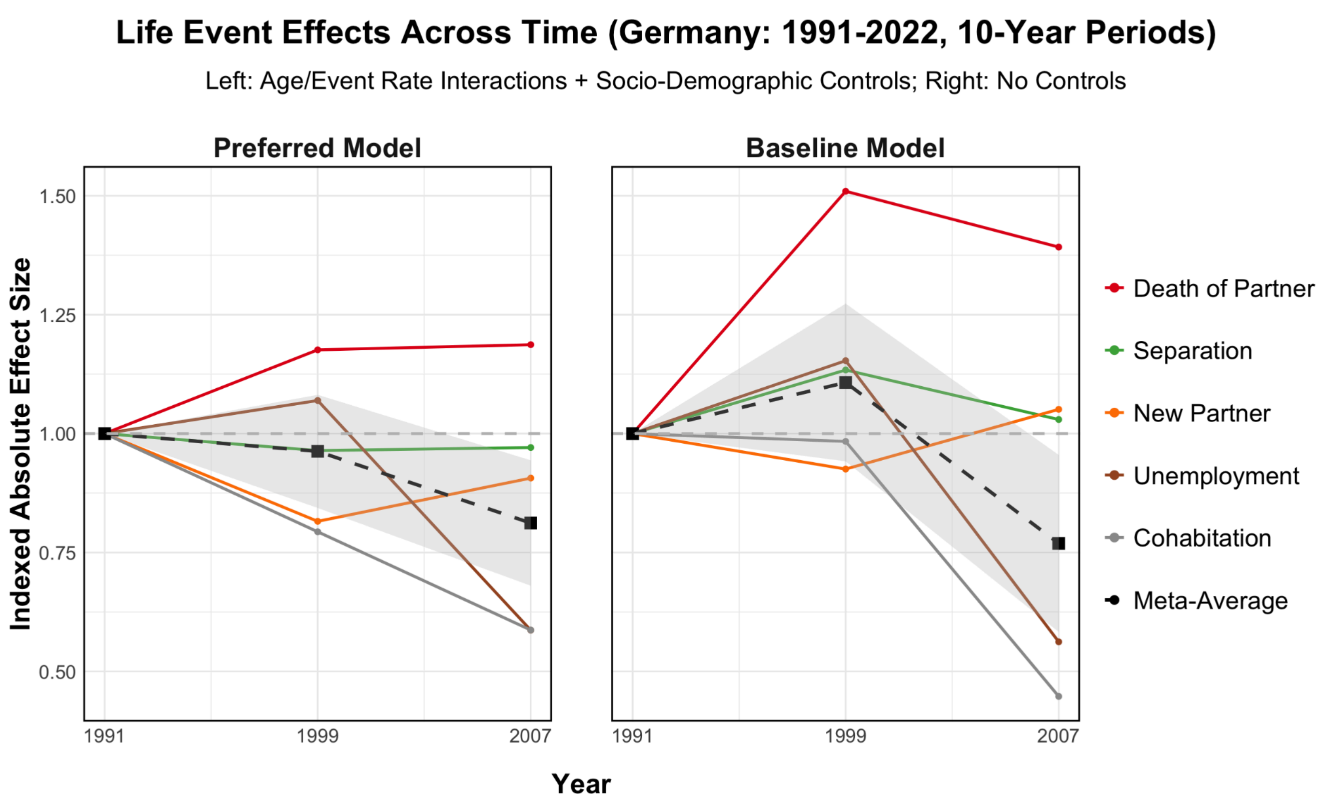
From equation (4), we can see that if there has been net positive shift, our reconstructed estimate (5) will underestimate the true value of h_i – for all individuals whose idea of a best and worst possible life are greater than zero ($h_t^{max} > h_t^{min} > 0$) from (1).

Appendix 2: Meta-Analyses with different specifications

Figure 11: Indexed Event Effect Sizes with Meta-Averages (3, 8, 10, and 15-year periods)




Life Event Effects Across Time (Germany: 1991-2022, 8-Year Periods)
Left: Age/Event Rate Interactions + Socio-Demographic Controls; Right: No Controls
Preferred Model
Baseline Model
Indexed Absolute Effect Size
1.6
1.2
0.8
0.4
1991
1999
2007
2015
Year
Death of Partner
Separation
Unemployment
New Partner
Cohabitation
Meta-Average


Life Event Effects Across Time (Germany: 1991-2022, 10-Year Periods)
Left: Age/Event Rate Interactions + Socio-Demographic Controls; Right: No Controls
Preferred Model
Baseline Model
Indexed Absolute Effect Size
1.50
1.25
1.00
0.75
0.50
1991
1999
2007
Year
Death of Partner
Separation
New Partner
Unemployment
Cohabitation
Meta-Average

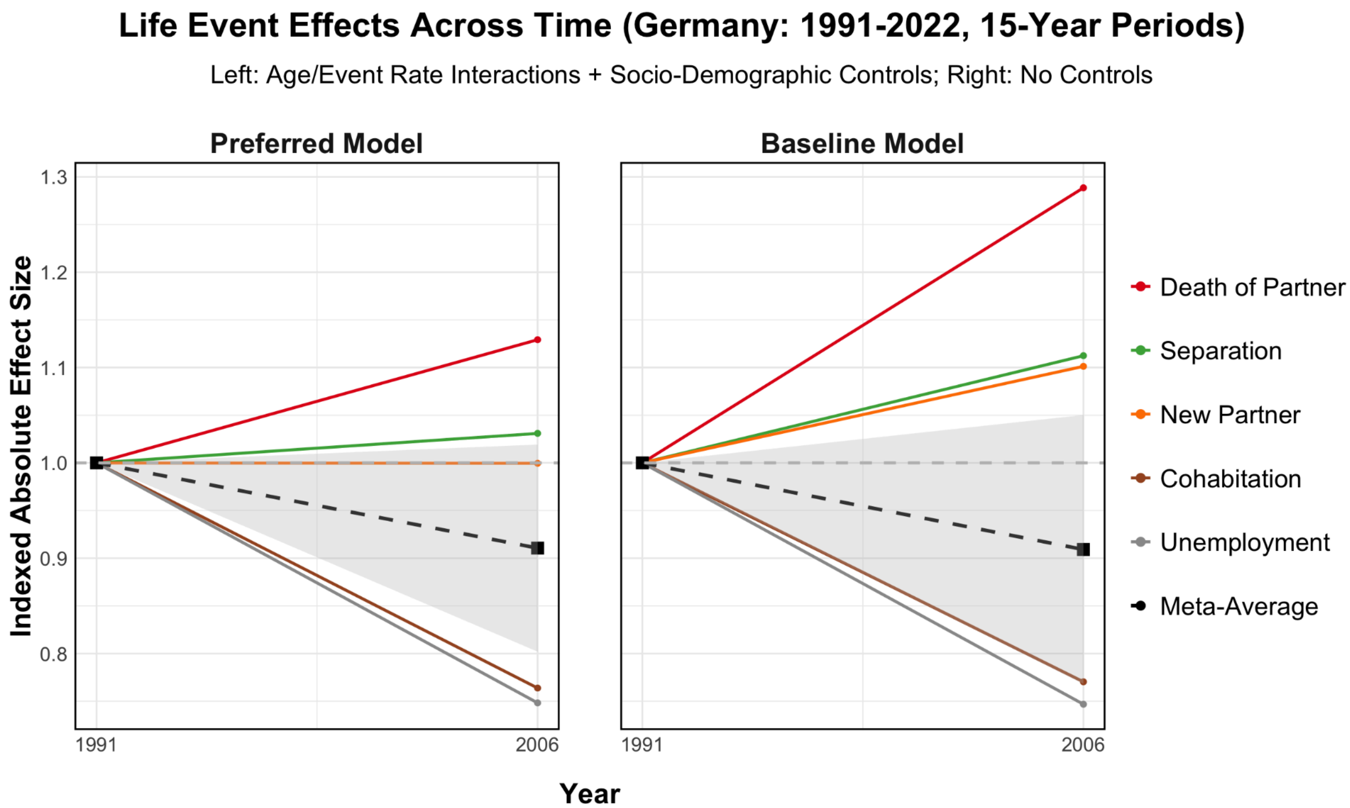

Life Event Effects Across Time (Germany: 1991-2022, 15-Year Periods)
Left: Age/Event Rate Interactions + Socio-Demographic Controls; Right: No Controls
Preferred Model
Baseline Model
Indexed Absolute Effect Size
1.3
1.2
1.1
1.0
0.9
0.8
1991
2006
Year
Death of Partner
Separation
New Partner
Cohabitation
Unemployment
Meta-Average

Figure 12: Comparing a Simple Average to the Meta-Average

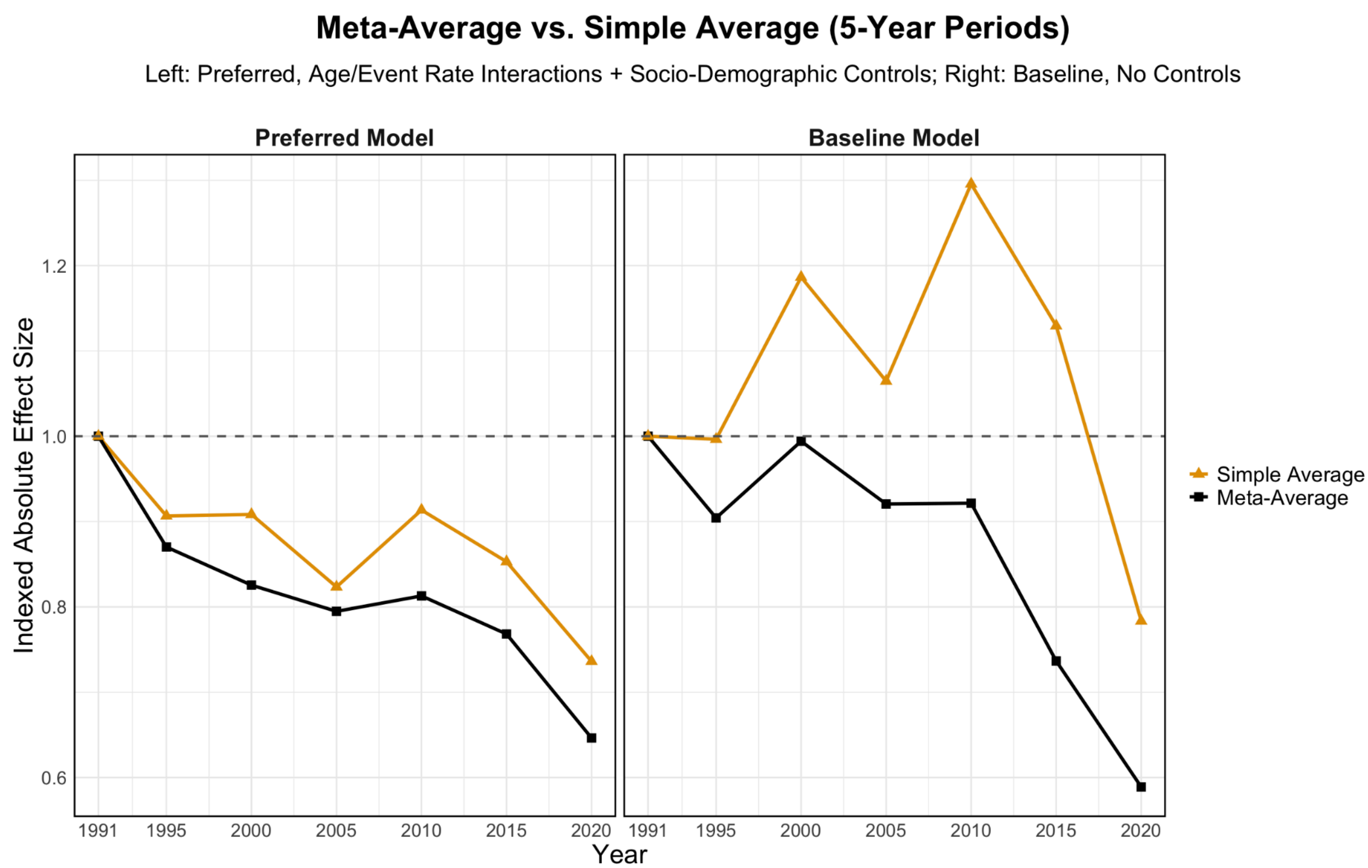


Figure 13: Alternative Model Specifications

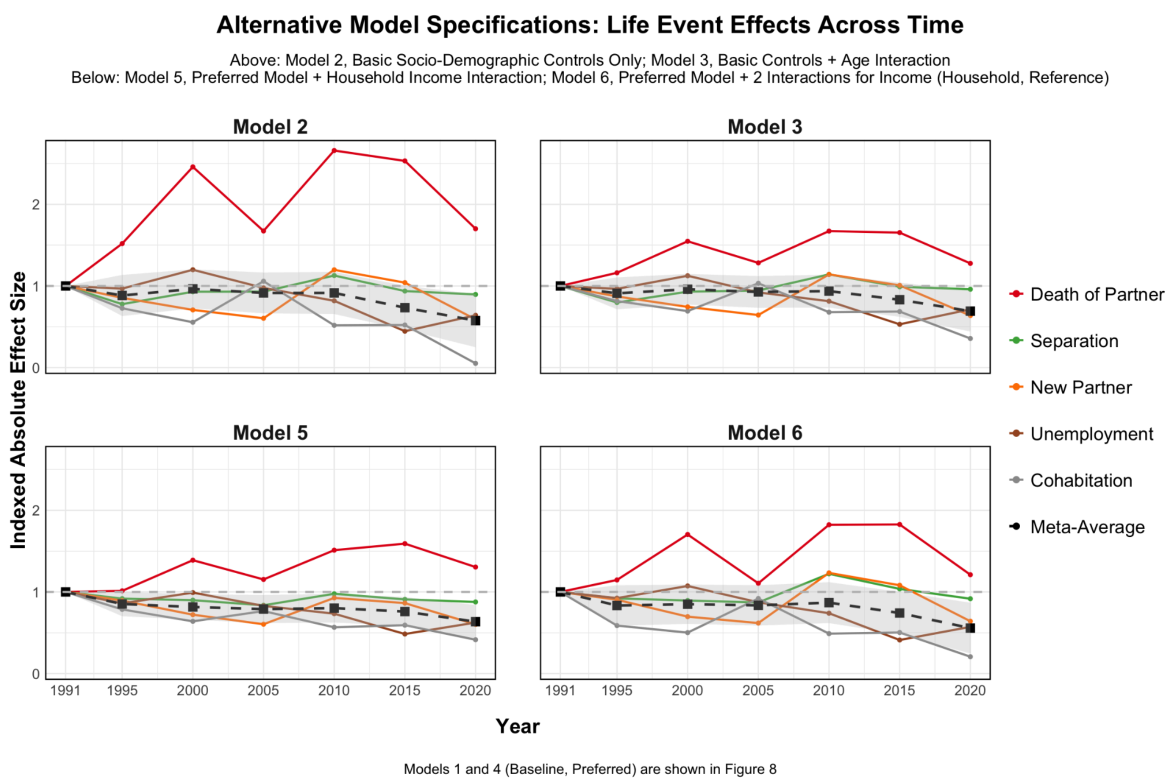
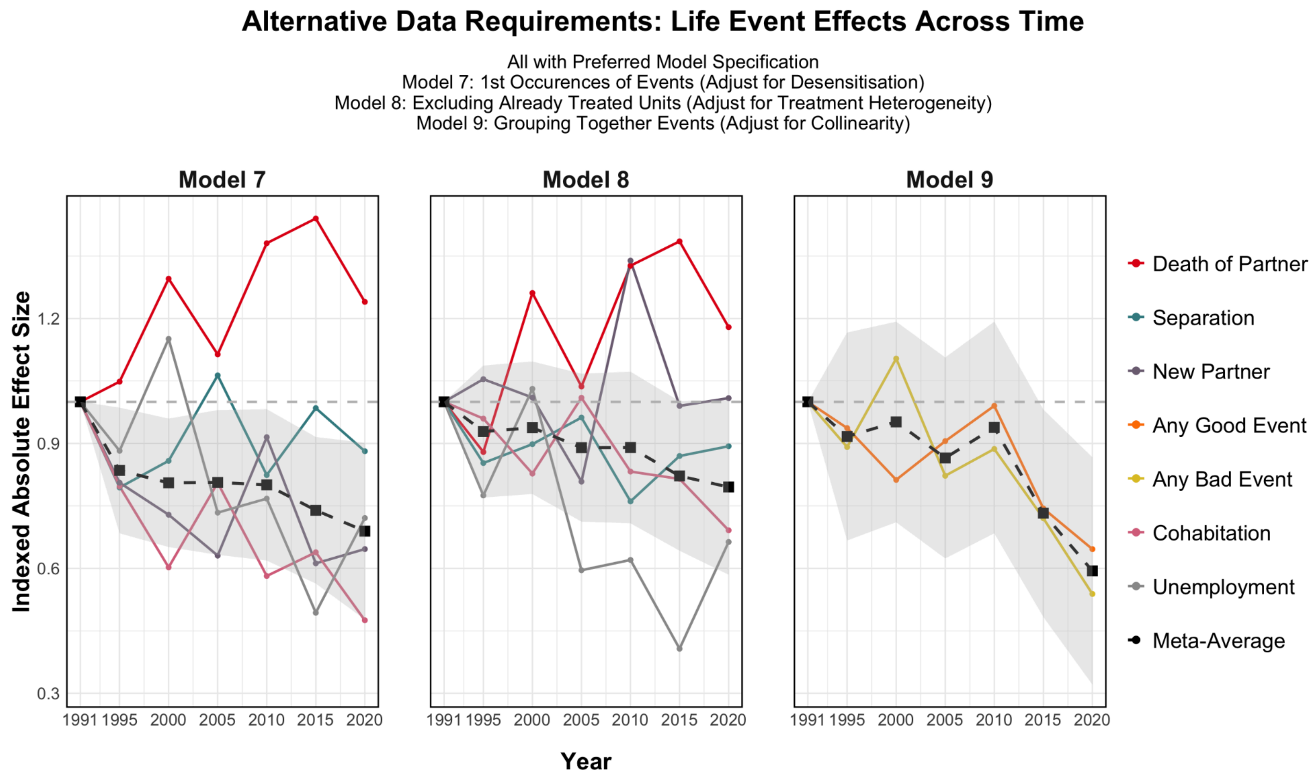


Figure 14: Alternative Data Requirements



Appendix 3: List of Controls

List of time-varying controls from Kaiser & Oswald (2022)

Controls used:

- Age – rendered in categories, rather than polynomials, for interpretative simplicity.
- Self-employment status
- Home ownership status
- Log of number of adults in the household
- Log of number of children in the household
- Childbirth indicator

- Marital status (categorical)
- Job status (categorical)
- Establishment size (categorical)
- Government region (categorical)

Controls excluded:

- Log of household income (personal and reference)
- Log of job hours
- Dummy for year survey (to avoid collinearity with period dummies)

Appendix 4: Inclusion Conditions for Treatment and Control Groups

Adapted from Krämer et al. (2025).

Event	Inclusion Criteria for Treatment	Inclusion Criteria for the Control:
Death of a Partner	Individual reports within annual questions on family changes.	Is in a partnership, and whose partner never died during the survey.
Separation	Biographical data on relationship spells	Never experienced a separation before entering the survey.
New Relationship	Biographical data on relationship spells	Never had a partnership before entering the survey.
Cohabitation	Biographical data on relationship spells	Never experienced cohabitation before entering the survey.
Unemployment	If an individual registered at the Employment Office, and were not registered as unemployed in the previous two years	Is still in the workforce (not already retired for at least 2 consecutive years aged 55 or older); is not already in long-term unemployment (not already registered as unemployed for at least 3 consecutive years); and never experienced unemployment during the survey.

Supplementary Material

Data-collection conducted in Stata (MP 18.5); meta-analysis and plotting via R (version 4.4.1). Code available on request from charlie.harrison.22@alumni.ucl.ac.uk

SOEP data is freely available for scientific use after signing a data contract. More information at: https://www.diw.de/en/soep.

Bibliography

Baker, A. C., Larcker, D. F., & Wang, C. C. Y. (2022). How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics, 144*(2), 370–395. https://doi.org/10.1016/j.jfineco.2022.01.004

Benjamin, D. J., Cooper, K., Heffetz, O., Kimball, M. S., & Zhou, J. (2023a). Adjusting for scale-use heterogeneity in self-reported well-being (NBER Working Paper No. 31728). National Bureau of Economic Research. https://doi.org/10.3386/w31728

Cinelli, C., Forney, A., & Pearl, J. (2024). A crash course in good and bad controls. *Sociological Methods & Research, 53*(3), 1071–1104. https://doi.org/10.1177/00491241221099552

Clark, A., Knabe, A., & Rätzel, S. (2009). *Boon or bane? Others' unemployment, well-being, and job insecurity* (IZA Discussion Paper No. 4210). Institute for the Study of Labor (IZA). https://doi.org/10.1016/j.labeco.2009.05.007

Clark, A. E., Diener, E., Georgellis, Y., & Lucas, R. E. (2008). Lags and leads in life satisfaction: A test of the baseline hypothesis. *The Economic Journal, 118*(529), F222–F243. https://doi.org/10.1111/j.1468-0297.2008.02150.x

Clark, A. E., Flèche, S., Layard, R., Powdthavee, N., & Ward, G. (2018). *The origins of happiness: The science of well-being over the life course.* Princeton University Press.

Clark, A. E., Frijters, P., & 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. https://doi.org/10.1257/jel.46.1.95

Deaton, A. (2011). Comment on "Work disability, work, and justification bias in Europe and the U.S." In D. A. Wise (Ed.), *Explorations in the economics of aging* (pp. 312–314). Chicago, IL: University of Chicago Press. https://doi.org/10.7208/chicago/9780226903385.001.0001

Easterlin, R. A. (1974). Does economic growth improve the human lot? Some empirical evidence. In P. A. David & M. W. Reder (Eds.), *Nations and households in economic growth* (pp. 89–125). Academic Press. https://doi.org/10.1016/B978-0-12-205050-3.50008-7

Fabian, M. (2022a). Scale norming undermines the use of life satisfaction scale data for welfare analysis. *Journal of Happiness Studies, 23*, 1509–1541. https://doi.org/10.1007/s10902-021-00460-8

Fabian, M. (2022b). Measuring subjective well-being. In *A theory of subjective well-being*. Oxford Academic. https://doi.org/10.1093/oso/9780197635261.003.0011

Fleurbaey, M., & Blanchet, D. (2013). *Beyond GDP: Measuring welfare and assessing sustainability*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199767199.001.0001

Frijters, P., Johnston, D. W., & Shields, M. A. (2011). Life satisfaction dynamics with quarterly life event data. *Scandinavian Journal of Economics, 113*, 190–211. https://doi.org/10.1111/j.1467-9442.2010.01638.x

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics, 225*(2), 254–277. https://doi.org/10.1016/j.jeconom.2021.03.014

Ghose, U., Krämer, M. D., Richter, D., Wagner, G. G., Infurna, F. J., Ram, N., & Gerstorf, D. (2025). How spousal bereavement shapes life satisfaction: Stability and change across historical time. *European Journal of Personality*. Advance online publication. https://doi.org/10.1177/08902070251336534

Graham, C., Eggers, A., & Sukhtankar, S. (2004). Does happiness pay? An exploration based on panel data from Russia. *Journal of Economic Behavior & Organization, 55*(3), 319–342. https://doi.org/10.1016/j.jebo.2003.09.002

Headey, B., & Muffels, R. (2018). A theory of life satisfaction dynamics: Stability, change, and volatility in 25-year life trajectories in Germany. *Social Indicators Research, 140*(2), 837–866. https://doi.org/10.1007/s11205-017-1785-z

Helliwell, J. F., Barrington-Leigh, C., Harris, A., & Huang, H. (2010). International evidence on the social context of well-being. In E. Diener, J. F. Helliwell, & D. Kahneman (Eds.), *International differences in well-being* (pp. 291–327). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199732739.003.0010

Howard, G. S., Ralph, K. M., Gulanick, N. A., Maxwell, S. E., Nance, D. W., & Gerber, S. K. (1979). Internal invalidity in pretest-posttest self-report evaluations and a re-evaluation of retrospective pretests. *Applied Psychological Measurement, 3*(1), 1–23. https://doi.org/10.1177/014662167900300101

Ingelström, M., & van der Deijl, W. (2021). Can happiness measures be calibrated? *Synthese, 199*, 5719–5746. https://doi.org/10.1007/s11229-021-03043-5

Kahneman, D., & Krueger, A. B. (2006). Developments in the measurement of subjective well-being. *Journal of Economic Perspectives, 20*(1), 3–24. https://doi.org/10.1257/089533006776526030

Kaiser, C., & Oswald, A. J. (2022). The scientific value of numerical measures of human feelings. *Proceedings of the National Academy of Sciences, 119*(42), e2210412119. https://doi.org/10.1073/pnas.2210412119

Kaiser, C. (2022). Using memories to assess the intrapersonal comparability of well-being reports. *Journal of Economic Behavior & Organization, 193,* 410–442. https://doi.org/10.1016/j.jebo.2021.11.009

Kettlewell, N., Morris, R. W., Ho, N., Cobb-Clark, D. A., Cripps, S., & Glozier, N. (2020). The differential impact of major life events on cognitive and affective wellbeing. *SSM - Population Health, 10,* 100533. https://doi.org/10.1016/j.ssmph.2019.100533

Krämer, M. D., Rohrer, J. M., Lucas, R. E., & Richter, D. (2025). Life events and life satisfaction: Estimating effects of multiple life events in combined models. *European Journal of Personality, 39*(1), 3–23. https://doi.org/10.1177/08902070241231017

Krueger, A. B., & Schkade, D. A. (2008). The reliability of subjective well-being measures. *Journal of Public Economics, 92*(8–9), 1833–1845. https://doi.org/10.1016/j.jpubeco.2007.12.015

Kubiszewski, I., Zakariyya, N., Costanza, R., & Jarvis, D. (2020). Resilience of self-reported life satisfaction: A case study of who conforms to set-point theory in Australia. *PLOS ONE, 15*(8), e0237161. https://doi.org/10.1371/journal.pone.0237161

Lipsey, M. W., & Wilson, D. B. (2001). Practical meta-analysis. Sage Publications, Inc.

Luhmann, M., & Eid, M. (2009). Does it really feel the same? Changes in life satisfaction following repeated life events. *Journal of Personality and Social Psychology, 97*(2), 363–381. https://doi.org/10.1037/a0015809

Malani, A., & Reif, J. (2015). Interpreting pre-trends as anticipation: Impact on estimated treatment effects from tort reform. *Journal of Public Economics, 124,* 1–17. https://doi.org/10.1016/j.jpubeco.2015.01.001

Montgomery, M. (2022). Reversing the gender gap in happiness. *Journal of Economic Behavior & Organization, 196,*65–78. https://doi.org/10.1016/j.jebo.2022.01.006

Oswald, A. J. (2008). On the curvature of the reporting function from objective reality to subjective feelings. *Economics Letters, 100*(3), 369–372. https://doi.org/10.1016/j.econlet.2008.02.032

Peichl, A., & Ungerer, M. (2017). Equality of opportunity: East vs. West Germany. *Bulletin of Economic Research, 69*(4), 421–427. https://doi.org/10.1111/boer.12103

Pinker, S. (2018). *Enlightenment now: The case for reason, science, humanism, and progress.* Viking.

Plant, M. (2020). A happy possibility about happiness (and other subjective) scales: An investigation and tentative defence of the cardinality thesis. *Happier Lives Institute.* https://www.happierlivesinstitute.org/report/the-comparability-of-subjective-scales/

Plant, M. (2024). A happy probability about happiness (and other) scales: An exploration and tentative defence of the cardinality assumption. *Wellbeing Research Centre.* doi.org/10.5287/ora-r1m7nj0kk

Powdthavee, N. (2007). Are there geographical variations in the psychological cost of unemployment in South Africa? *Social Indicators Research, 80*(3), 629–652. https://doi.org/10.1007/s11205-006-0013-z

Prati, A., & Senik, C. (2022). Feeling good is feeling better. *Psychological Science, 33*(11), 1828–1841. https://doi.org/10.1177/09567976221096158

Prati, A., & Senik, C. (2025). Is it possible to raise national happiness? *CEP Discussion Paper No. 2068.* Centre for Economic Performance, London School of Economics and Political Science. https://cep.lse.ac.uk/pubs/download/dp2068.pdf

Ravallion, M., & Lokshin, M. (2001). Identifying welfare effects from subjective questions. *Economica, 68*, 335–357. https://doi.org/10.1111/1468-0335.00250

Rosling, H. (2018). *Factfulness: Ten reasons we're wrong about the world—and why things are better than you think.* Flatiron Books.

Schutte, L., Negri, L., & Delle Fave, A. (2021). Rasch analysis of the Satisfaction with Life Scale across countries: Findings from South Africa and Italy. *Current Psychology, 40,* 4908–4917. https://doi.org/10.1007/s12144-019-00424-5

Schwartz, C. E., Stucky, B., Rivers, C. S., Noonan, V. K., Finkelstein, J. A., & RHSCIR Network.(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.e1. https://doi.org/10.1016/j.apmr.2018.01.028

Schwarz, N., & Strack, F. (1999). Reports of subjective well-being: Judgmental processes and their methodological implications. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 61–84). Russell Sage Foundation.

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics, 106*(4), 1039–1061. https://doi.org/10.2307/2937956

van Praag, B. M. S. (1991). Ordinal and cardinal utility: An integration of the two dimensions of the welfare concept. *Journal of Econometrics, 50*(1–2), 69–89. https://doi.org/10.1016/0304-4076(91)90090-Z

Wadsworth, T. (2016). Marriage and subjective well-being: How and why context matters. *Social Indicators Research, 126*, 1025–1048. https://doi.org/10.1007/s11205-015-0930-9

Whyte, J. (2013). *Quack policy: Abusing science in the cause of paternalism*. Institute of Economic Affairs. https://doi.org/10.2139/ssrn.3915620

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). MIT Press.

Wooldridge, J. M. (2013). *Introductory econometrics: A modern approach* (5th ed.). South-Western, Cengage Learning.