Published in final edited form as:

J Res Pers. 2007 October ; 41(5): 1091–1098.

How Stable is Happiness? Using the STARTS Model to Estimate the Stability of Life Satisfaction

Richard E. Lucas

Michigan State University & German Institute for Economic Research

M. Brent Donnellan

Michigan State University

Abstract

A common interpretation of existing subjective well-being research is that long-term levels of well-being are almost completely stable. However, few studies have estimated stability and change using appropriate statistical models that can precisely address this question. The STARTS model (Kenny & Zautra, 2001) was used to analyze life satisfaction data from two nationally representative panel studies. Results show that 34–38% of the variance in observed scores is trait variance that does not change. An additional 29–34% can be accounted for by an autoregressive trait that is only moderately stable over time. Thus, although life satisfaction is moderately stable over long periods of time, there is also an appreciable degree of instability that might depend on contextual circumstances.

Keywords

Subjective Well-Being; Life Satisfaction; Stability; STARTS Model

Subjective well-being (SWB) reflects the extent to which people think and feel that their life is going well. Because SWB is strongly desired by people from around the world (Diener, 2000), psychologists strive to identify factors that can foster high levels of well-being. However, this desire is predicated on the assumption that SWB is actually responsive to changing life circumstances. If SWB is a completely stable individual difference that is independent of contextual factors, then there would be little that could be done to promote lasting change. Instead, people would be stuck on a “hedonic treadmill” (Brickman & Campbell, 1971). Although their well-being might change in response to life events, it would inevitably return to a genetically determined set-point after a period of adaptation.

Indeed, the existing evidence suggests that the effects of external circumstances on happiness are relatively small. For instance, demographic characteristics and life circumstances do not account for large amounts of variance in well-being (Argyle, 1999). In contrast, stable, genetically-influenced personality characteristics exhibit moderate to strong correlations with

Address correspondence to: Richard E. Lucas Department of Psychology Michigan State University East Lansing, MI 48823 E-Mail: lucasri@msu.edu Phone: (517) 432–4360 Fax: (517) 432–2476.

The data used in this paper were made available by the German Institute for Economic Research and by the UK Data Archive. The BHPS data were originally collected by the ESRC Research Centre on Micro-social Change at the University of Essex, now incorporated within the Institute for Social and Economic Research. Neither the original collectors of the data nor the Archive bear any responsibility for the analyses or interpretations presented here. Richard E. Lucas and M. Brent Donnellan were supported by National Institute on Aging grants 1R03AG026028–01 and 1R03AG028744–01 during the preparation of this manuscript.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

well-being constructs (see Lucas, in press, for a review). These findings have led some to suggest that that long-term levels of happiness cannot change and that people can adapt to almost any life circumstance including severe disabilities and economic windfalls.

In one of the strongest statements against the possibility for change, Lykken and Tellegen (1996) suggested that “trying to be happier [may be] as futile as trying to be taller” (p. 189). These authors based this statement on evidence from behavioral genetic studies. Although it is widely known that the various components of SWB are heritable, Lykken and Tellegen extended the standard twin design by assessing twins' SWB at two time points separated by approximately 10 years. Consistent with previous studies, they found that the heritability at the initial assessment was approximately 50%. But more importantly, they showed that about 80% of the long-term stable variance was heritable. These results suggest that although external factors may play a role in short-term fluctuations, long-term levels of SWB are primarily affected by genetically determined characteristics that are currently beyond control.

Of course, the importance of this finding depends on the extent to which variance in happiness is in fact stable. If only a small proportion of the variance in well-being measures is stable over time, then the fact that this stable portion is almost exclusively related to genetic factors is of academic interest but has few implications for people's actual experience. Lykken and Tellegen (1996) estimated the 10-year stability to be about .50 in their sample, which would mean that 50% of the variance at either occasion was stable trait variance. However, this figure might represent an upper bound estimate as more recent work has indicated that the stability of well-being varies across samples and measures. Schimmack and Oishi (2005) conducted a meta-analysis of stability estimates from life satisfaction scales, and they showed that the average 10-year stability is approximately .35. Similarly, Fujita and Diener (2005) used data from the German Socio-Economic Panel Study (GSOEP), an ongoing, nationally representative panel study of people living in Germany, to estimate the stability of a single-item life satisfaction measure. Like Schimmack and Oishi, they found 10-year stability estimates of around .35.

Although correlations between measures of well-being assessed at two points in time provide broad clues about the stability of SWB over particular intervals, they provide an incomplete picture of stability and change. This is because similar correlations between two measures could result from a variety of underlying processes (Fraley & Roberts, 2005). For instance, a strong one-year stability coefficient could result in very little stability over the long-term if SWB had a simple first-order autoregressive structure (where an individual's score depended only on his or her previous score plus a random component). For example, if the one-year stability was .85 and life satisfaction conformed to an autoregressive structure, then the predicted stability over 20 years would be just .04. In contrast, if life satisfaction was best described using a stable-trait model (where an individual's score depended on his or her unchanging trait score plus a random component that varied from year to year), then even moderate one-year stabilities could be consistent with a relatively high degree of stability over the long-term. Both Schimmack and Oishi (2005) and Fujita and Diener (2005) found evidence that stable trait variance plays some role in the stability of life satisfaction measures. However, multi-wave data and appropriate structural models are necessary to establish just how much variance a stable underlying trait contributes to repeated measures of life satisfaction.

In an earlier investigation of this issue, Ehrhardt, Saris, and Veenhoven (2000) analyzed 11 waves of life satisfaction data from the same panel study that Fujita and Diener (2005) examined. Their results showed that 29% of the variance was accounted for by a stable trait, 29% by an autoregressive trait, and 42% by a combined state/error component. However, there are a number of limitations of this earlier investigation. First, like Fujita and Diener, the authors used listwise deletion of participants with missing data. Because considerable attrition occurs in longitudinal studies, this strategy often results in the exclusion of many participants from

the analysis. Put simply, listwise deletion can result in biased estimates and is therefore not an optimal strategy for dealing with missing data (Schafer & Graham, 2002).

Second, although 11 years may seem like a long time for a longitudinal study, it is possible that even longer time frames are needed to separate stable trait variance from autoregressive trait variance. Stable trait models imply that correlations will eventually reach some minimum level beyond which they do not drop (i.e., they asymptote), even over increasingly long periods of time. If studies are not conducted for a sufficient period of time this minimum may never be reached (see e.g., Figure 1 in Fraley & Roberts, 2005, p. 61), and a simpler autoregressive model with no stable trait (or a relatively small stable trait) may appear to describe the data as well as more complicated models that incorporate stable trait factors. In the current studies, we attempt to replicate Ehrhardt et al.'s (2000) findings using full information maximum likelihood estimation and 10 additional years of data from the GSOEP.¹ We then replicate these analyses from a separate nationally representative longitudinal study, the British Household Panel Study (BHPS). Together, these studies should provide a much more nuanced understanding of the stability of SWB over time.

Method

Sample 1: The German Socio-Economic Panel Study (GSOEP)

The data from this sample come from Waves 1 through 21 of the GSOEP, a longitudinal study of German households that began in 1984 (see Haiken-De New & Frick, 2005, for details). All respondents who participated in the first wave of data collection were included in the analyses, regardless of how many more waves of data they contributed. A total of 8,632 respondents who participated for an average of 11.75 years ($sd = 7.86$) were included. Each year, participants were asked a single question about how satisfied they were with their lives on scale that ranged from 0 to 10.

Sample 2: The British Household Panel Study (BHPS)

The data for this sample come from the BHPS, a longitudinal study of British households that began in 1991 (University of Essex, Institute for Social and Economic Research, 2004). Because the life satisfaction measure was only administered in eight of the nine most recent waves, only data from these waves are included. All respondents who participated in the first wave during which life satisfaction was assessed were included in the analyses, regardless of how many more waves of data they contributed. In total, 9,437 respondents participated for an average of 7.36 years ($sd = 2.65$). The life satisfaction item read "How dissatisfied or satisfied are you with your life overall," and participants answered using a scale that ranged from 1 to 7.

Analytic Technique

A number of models for decomposing multi-wave data have been proposed (see Cole, Martin, & Steiger, 2005, for a review). However, many of these models require multiple indicators at each measurement occasion and therefore cannot be used to assess the stability of single-item life satisfaction measures that are the norm among nationally representative panel studies. One exception, however, is the STARTS model proposed by Kenny and Zautra (2001). This model

¹Schilling (2006) also investigated the longitudinal structure of life satisfaction using the GSOEP data. However, he restricted his analyses to a pure first-order autoregressive model. Such a model would imply that the correlation between two non-contiguous waves would be equal to the one-year stability raised to the n^{th} power, where n is equal to the number of intervening waves. For example, if the one-year stability is .50, then this implies that the three-year stability is $(.50)^3$ or .125. However, if the longitudinal data have a latent-trait structure (as our analyses show they do), then the long-term stabilities will be higher than expected given the one-year stabilities. In these cases, a simple autoregressive model will artificially inflate estimates of both the year-to-year stabilities and the wave-specific residual component to account for the relatively high long-term stabilities.

decomposes the variance in each wave into three latent factors or components: (a) a stable trait component that does not change (labeled ST); (b) an autoregressive component that depends on a person's standing during the previous occasion (labeled ART); and, (c) a state component that includes both occasion specific variance and measurement error (labeled S).

The basic formulation of the STARTS model involves several constraints (Kenny & Zautra, 2000). For instance, in the canonical STARTS model, the ART and S variances are constrained to the same value across all measurement occasions. Similarly, the path coefficients linking the autoregressive traits across occasions are constrained to the same value. These constraints facilitate model identification and produce standardized stability coefficients for the AR traits. However, there may be cases where the assumptions underlying these constraints are not tenable. For example, if there are panel conditioning effects and participants use the scale differently after repeated assessments, then variance in the S and ART components may change over time. Similarly, the assumption that stability coefficients do not change might be problematic in developmental studies where there is evidence that individual differences get increasingly stable with age (e.g. Roberts & DelVecchio, 2000).

Given these complexities and the fact that we have enough waves of data to freely estimate most model parameters, we do not impose these constraints. Instead, we only assume that the amount of variance accounted for by the state component remains constant across waves. By freeing these constraints, it is not possible to calculate an omnibus variance decomposition because the AR variances may fluctuate across occasions leading to different amounts of total variance at each wave. Thus, the percent of stable trait, autoregressive trait, and state variance will also change across waves. Therefore, we examine how much change occurs in these parameters across waves and report the average values of ST, ART, and S variance.

To determine whether the variance decomposition changes with age, we estimate separate models for different cohorts using multi-group modeling techniques. Specifically, we estimated models for six groups: those who were under 26, 26–30, 31–40, 41–50, 51–60 and over 60 at the start of the survey. If life satisfaction becomes more stable with age, the percentage of variance accounted for by the stable trait component should increase across cohorts.

Results

We evaluated model fit using the χ^2 exact-fit test, the root mean square error of approximation (RMSEA), the Incremental Fit Index (IFI), and the Tucker-Lewis Index (TLI). Because the sample sizes are quite large, the exact-fit test will likely be significant even when the model is trivially mis-specified. Therefore, we consider these models to be acceptable if the RMSEA value is lower than .05 and the IFI and TLI values are close to 1.00.

We first tested a model using participants from the GSOEP. Although this model did not pass the χ^2 exact-fit test, the other fit indices indicated that the model described the data well: $\chi^2 = 548.00$, $df = 188$; RMSEA = .013; IFI = 0.99; TLI = 0.99. On average, the stable trait component accounted for 34% of the observed variance in any single wave of assessment, an estimate that is slightly higher than that found by Ehrhardt et al. (2000). In contrast, our analyses resulted in lower estimates of unstable state/error variance than did Ehrhardt et al.'s analyses (33% versus 42%). This value is important because it provides a lower bound for the reliability of this single-item measure. If 33% of the variance can be accounted for by a combined state/error component, then this means that the reliability of this measure is at least .67.

The final component—the autoregressive trait—accounted for an average of 34% of the observed variance in any single wave. This component reflects the part of life satisfaction that is stable in the short term but that can fluctuate over time. On average, the standardized

regression coefficient linking two subsequent waves was .75, which suggests that even this changeable part of life satisfaction leads to considerable stability in the short term. Notably, the stability of the autoregressive component increased across waves from .40 in the first two waves of the study to .91 from Waves 20 to 21. This trend could result from developmental change or from testing effects that emerge when people answer the same questions repeatedly over time. The open squares in Figure 1 show the predicted stability from 1 to 20 years, based on the estimates from this model. Stability decreases at first but then levels off with increasing time. Even over a 21-year period, however, stability is predicted to be approximately .30.

These basic findings replicated in the BHPS. The unconstrained STARTS model fit the data well, passing the exact-fit test and exceeding the criteria for the other indices: $\chi^2 = 29.69$, $df = 19$, ns ; RMSEA = .006; IFI = 1.00; TLI = 0.99. The stable trait component accounted for an average of 38% of the observed variance, whereas the unstable state/error component accounted for an average of 33%. Again, these estimates suggest that the reliability of this single-item measure is at least .67, and that one-third of the total variance is trait variance that is completely stable over considerable periods of time. Also consistent with estimates from the GSOEP sample, 29% of the variance could be accounted for by an autoregressive trait that had an average standardized stability coefficient of .76. Unlike the estimates from the GSOEP sample, however, the stability coefficients for the autoregressive component tended to increase only slightly over time. The predicted stabilities based are presented as solid diamonds in Figure 1.

As a final examination of the stability of life satisfaction, we conducted multi-group analyses in each of the two samples to determine whether the variance decomposition varied across cohorts. Neither model passed the exact-fit test, but the other indices suggested that both models described the data well: $\chi^2 = 1913.79$, $df = 1131$; RMSEA = .009; IFI = 0.98; TLI = 0.98 in the GSOEP and $\chi^2 = 211.38$, $df = 115$; RMSEA = .009; IFI = 0.99; TLI = 0.99 in the BHPS.² As Figure 2 shows, the amount of variance that could be accounted for by the stable trait component was lowest among those under age 26 and highest for those between 51 and 60. Notably, in both the GSOEP and BHPS samples, the percentage of variance accounted for by this stable trait dropped slightly in the oldest group.

Discussion

Behavioral genetic studies suggest that the stable component of SWB is strongly heritable. One common interpretation of this finding is that changing a person's happiness is nearly impossible. However, Rutter (1997) noted that estimates of heritability “provide no unambiguous implications for theory, policy, or practice” (p. 391). In addition, it is unclear precisely how much of the variability in SWB is completely stable over time. Indeed, it has been difficult to determine the extent to which a stable underlying trait contributes to variance in observed measures because many longitudinal studies have been conducted for short periods of time or have collected only a few waves of data. We sought to address this gap by separating stable trait variance from autoregressive trait variance and unstable state variance in life satisfaction data from two long-running, nationally representative panel studies.

Estimates were remarkably consistent across the two samples. The stable trait component accounted for between 34 and 38% of the variance in life satisfaction at any given wave. This trait reflects the part of life satisfaction that exhibits complete stability over time, and it is this

²In a few cases in the latter waves, the residual variance for the autoregressive component in some groups was estimated to be less than zero. These inadmissible values likely result from the fact that these variances approach zero as the autoregressive component becomes more and more stable over time. Because it was reasonable to assume that these variances were in fact zero, we constrained them to be zero. In addition, as a second approach, we used Bayesian estimation, setting prior probabilities for variances less than zero equal to zero. The means of the posterior distributions were very similar to the estimates from the models where variances were constrained to be zero.

component that is typically the source of generalizations about the futility of changing happiness. In addition, a moderately stable autoregressive component accounts for between 29 and 34% of the variance in repeated assessments of SWB. This component may reflect the influence of lasting but modifiable life circumstances on SWB. For instance, previous research using the GSOEP showed that widows and widowers experience strong reactions to the loss of a spouse, but that this reaction subsides over time (Lucas, Clark, Georgellis, & Diener, 2003). Such significant life events could lead to short-term stability that decreases over time as people adapt.

Finally, the multi-group analyses from both samples showed that the variance decomposition changed across cohorts. In particular, the stable trait component played an increasingly important role in life satisfaction judgments with age. This trend towards increasing stability is consistent with existing work on other individual differences (Roberts & DelVecchio, 2000). Moreover, the source of this stability could be due to the fact that life circumstances become more stable as people age. Together, these results show that although life satisfaction measures are very stable over the short term and moderately stable over the long-term, there is room for change. As such, broad pronouncements that changing happiness is akin to changing height should probably be tempered.

References

- Argyle, M. Causes and correlates of happiness.. In: Kahneman, D.; Diener, E.; Schwarz, N., editors. *Well-being: The foundations of hedonic psychology*. Russell Sage Foundation; New York, NY: 1999. p. 353-373.
- Cole DA, Martin NC, Steiger JH. Empirical and conceptual problems with longitudinal trait-state models: Introducing a Trait-State-Occasion Model. *Psychological Methods* 2005;10:3–20. [PubMed: 15810866]
- Diener E. Subjective well-being: The science of happiness and a proposal for a national index. *American Psychologist* 2000;55:34–43. [PubMed: 11392863]
- Ehrhardt JJ, Saris WE, Veenhoven R. Stability of life-satisfaction over time: Analysis of change in ranks in a national population. *Journal of Happiness Studies* 2000;1:177–205.
- Fraley RC, Roberts BW. Patterns of continuity: A dynamic model for conceptualizing the stability of individual differences in psychological constructs across the life course. *Psychological Review* 2005;112:60–74. [PubMed: 15631588]
- Fujita F, Diener E. Life satisfaction set point: Stability and change. *Journal of Personality and Social Psychology* 2005;88:158–164. [PubMed: 15631581]
- Haisken-De New, JP.; Frick, R. Desktop companion to the German Socio-Economic Panel Study (GSOEP). German Institute for Economic Research (DIW); Berlin: 2005.
- Kenny, DA.; Zautra, A. The trait-state models for longitudinal data.. In: Collins, LM.; Sayer, AG., editors. *New Methods for the Analysis of Change*. American Psychological Association; Washington, DC: 2001. p. 243-263.
- Lucas, RE. Eid, M.; Larsen, RJ., editors. *Personality and subjective well-being.. The science of subjective well-being*. (In press)
- Lucas RE, Clark AE, Georgellis Y, Diener E. Reexamining adaptation and the set point model of happiness: Reactions to changes in marital status. *Journal of Personality & Social Psychology* 2003;84:527–539. [PubMed: 12635914]
- Lykken D, Tellegen A. Happiness is a stochastic phenomenon. *Psychological Science* 1996;7:186–189.
- Roberts BW, DelVecchio WF. The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin* 2000;126:3–25. [PubMed: 10668348]
- Rutter ML. Nature-nurture integration: The example of antisocial behavior. *American Psychologist* 1997;52:390–398.

- Schafer JL, Graham JW. Missing data: Our view of the state of the art. *Psychological Methods* 2002;7:147–177. [PubMed: 12090408]
- Schilling O. Development of life satisfaction in old age: Another view on the “paradox.”. *Social Indicators Research* 2006;75:241–271.
- Schimmack U, Oishi S. Chronically accessible versus temporarily accessible sources of life satisfaction judgments. *Journal of Personality and Social Psychology* 2005;89:395–406. [PubMed: 16248721]
- University of Essex. Institute for Social and Economic Research. British Household Panel Survey; Waves 1–12, 1991–2003. UK Data Archive; Colchester, Essex: 2004. [computer file][distributor]

Figure 1.
Predicted stability of life satisfaction in the GSOEP and BHPS

Figure 2.

Average amount of variance accounted for by the stable trait component across various age groups in the GSOEP and BHPS.