



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Headey, B;Muffels, R

Title:

Towards a Theory of Medium Term Life Satisfaction: Similar Results for Australia, Britain and Germany

Date:

2017-10-01

Citation:

Headey, B. & Muffels, R. (2017). Towards a Theory of Medium Term Life Satisfaction: Similar Results for Australia, Britain and Germany. *Social Indicators Research*, 134 (1), pp.359-384. <https://doi.org/10.1007/s11205-016-1430-2>.

Persistent Link:

<https://hdl.handle.net/11343/282584>

**TOWARDS A THEORY OF MEDIUM TERM LIFE
SATISFACTION: SIMILAR RESULTS FOR AUSTRALIA,
BRITAIN AND GERMANY***

* We would like to thank Ulrich Schimmack of the University of Toronto at Mississauga and Derek Headey of the International Food Policy Research Institute (IFPRI) for helpful suggestions on modelling reciprocal causation.

**TOWARDS A THEORY OF MEDIUM TERM LIFE
SATISFACTION: SIMILAR RESULTS FOR AUSTRALIA,
BRITAIN AND GERMANY**

Abstract

We analyse the *Life Satisfaction trajectories* of respondents in three long-running, national panel surveys: the Household, Income and Labour Dynamics Australia Survey (HILDA), the British Household Panel Survey (BHPS) and the German Socio-Economic Panel (SOEP). Previous research has shown that substantial minorities of respondents in all three countries recorded long term changes in LS (Fujita and Diener, 2005; Headey, 2006; Headey, Muffels and Wagner, 2010, 2013). In a recent SIR paper based on the German data (Headey and Muffels, 2015), we showed graphs of LS trajectories which suggested - and subsequent statistical analysis confirmed - that respondents typically spend multiple consecutive years above and, in other periods, below their own long term mean level of LS. Here we extend the analysis to Australia and Britain, showing that results replicate in two more Western countries. It appears that most people go through relatively happy periods of life, and relatively unhappy periods. The evidence runs counter to set-point theory which views adult LS as stable, except for short term fluctuations due to life events.

In the second half of the paper we try to contribute to a theory of medium term life satisfaction. We estimate structural equation models with two-way causation between LS and variables usually treated as causes of LS, including health, social support, frequency of social activities, and satisfaction with one's work, partner and family life. In all three countries we find that there are *positive feedback loops* between these variables and LS, which partly account for extended periods of high or low LS.

The two-way causation models are based on a modified concept of 'Granger-causation' (Granger, 1969). The main intuition behind Granger-causation is that if x can be shown to be statistically significantly related to y in a model which includes multiple lags of y , then it can be inferred that x is one cause of y .

Keywords: trajectories of life satisfaction; set-point theory; two-way causation; Granger-causation; positive feedback loops; medium term change

INTRODUCTION

A problem with many social science theories, including theories of happiness or life satisfaction, is that they attract criticism, sometimes derision, because they seem little more than common sense. They can seem like common sense even when they contradict each other. Take the set-point theory of happiness. This theory rests on the common-sense sounding claim that adult happiness is stable – it has a set-point - due to stable personality traits. Some people, the theory claims, are persistently happy because they rate low on neuroticism and high on extroversion. Others are faced with persistent unhappiness because of unhappy personalities; high neuroticism and low extroversion. In this paper we put forward an alternative, but contradictory claim, which may also sound like common sense. Our claim is that most people go through extended relatively happy periods of life, in which their happiness is above their own long term average rating, and also extended unhappy periods. This description probably accords with most people's recollections of their own lives to date, and so might also be dismissed as common sense.

Trajectories of Life Satisfaction

In order to establish whether adult life satisfaction (LS) is stable, or whether people typically experience extended periods in which their LS is

above or below their long term average, it is essential to access long term panel data. Three panel datasets which provide long term evidence about *trajectories of life satisfaction* (LS) are the Household, Income and Labour Dynamics Australia Survey (HILDA), the British Household Panel Survey (BHPS) and the German Socio-Economic Panel (SOEP). We analyse the trajectories of individuals who recorded their LS for twelve consecutive years in the Australian panel (2001-12), thirteen years in the British panel (1996-2008), and twenty years in the German panel (1991-2010). Graphs of LS trajectories, and ensuing statistical analysis, show that for many individuals LS is highly volatile. Further, it will become clear that many spend extended periods of time – multiple consecutive years - in which their LS is either well above or well below their own long-term mean.

In previous papers it has been shown that substantial minorities in each country record long-term changes in LS (Fujita and Diener, 2005; Headey, 2006; Headey, Muffels and Wagner, 2010, 2013). Our results here on LS trajectories provide additional evidence that the *set-point theory of LS* (Brickman and Campbell, 1971; Lykken and Tellegen, 1996) can no longer be maintained. Set-point theory views adult LS as stable, subject only to short term fluctuations due to life events. As most contributors to a recent collection of articles on *The Stability of*

Happiness (Sheldon and Lucas, 2014) concluded, the main challenge for researchers now is to develop a theory which accounts for medium and long term change.

In this paper we suggest some elements of a theory of change. We reassess associations between LS and quite a wide range of variables usually treated as causes of LS: health, exercise, social support, frequency of socialising with friends, relatives and neighbours, and satisfaction with one's work, partner and family life. In review articles it has been suggested these and many other variables could be consequences of LS, or perhaps both causes and consequences (Diener, 1984; Diener, Suh, Lucas and Smith, 1999; Frey and Stutzer, 2002). Nevertheless, the possibility of two-way causation is routinely ignored in empirical research. In the case of the link between LS and health, this must be a mistake. Prospective studies have shown that individuals who rate high on LS at baseline live longer than those with lower initial LS ratings (Deeg and van Zonneveld, 1989; Headey, Hoehne and Wagner, 2014). This could only happen if happiness promotes better health. Previous evidence is not so convincing in relation to other possible two-way links. However, it seems plausible to hypothesise that happiness could cause as well as be caused by frequency of exercise, social support, social participation and satisfaction with work and family life.

Two-way causation – *positive feedback loops* – appears partly to account for multi-year periods of high or low LS. For example, changes in frequency of exercise, occurring in year t , are estimated to produce changes in LS in year $t+1$. Then, in a positive feedback loop, changes that have occurred in LS are positively associated with changes in exercise in year $t+2$, which in turn are associated with LS at $t+3$.

Cumulative evidence against set-point theory

In the last fifteen years or so, set-point theory (Brickman and Campbell, 1971; Lykken and Tellegen, 1996) has come under increasing challenge. Easterlin (2003), whose previous research on the adaptation of LS to gains and losses in the financial domain had lent support to the theory, reviewed the effects of changes in non-economic domains, mainly health and family life, and concluded that in these domains adaptation to change is partial, but far from complete. He reported that individuals who develop chronic health problems record, on average, permanently lowered LS (Easterlin, 2003; Mehnert, Kraus, Nadler and Boyd, 1990) A substantial proportion of individuals who get married report long term gains in LS (Lucas, Clark, Georgellis and Diener, 2003). It has also been shown that individuals who experience unemployment more than once

record prolonged declines in LS, and that full recovery never occurs in many cases (Luhmann and Eid, 2009; Luhmann, Hoffman, Eid and Lucas, 2012).

The limitations of set-point theory became clearer as long term panel data became available. Fujita and Diener (2005) analysed data from the first seventeen years of SOEP and found that many respondents recorded substantial changes in LS. Headey (2006) and Headey, Muffels and Wagner (2010, 2013), using data from the Australian and British household panels as well as SOEP, found that a quarter to a third of survey respondents recorded substantial gains or losses of LS over periods ranging from a decade to twenty-five years.

Several studies by positive psychologists have provided partial explanations of change in LS. Experimental evidence has indicated the beneficial effects on LS of altruistic actions, religious practice, and activities which play to one's character strengths (Diener and Seligman, 2004; Lyubomirsky, Sheldon and Schkade, 2005; Dunn, Aknin and Norton, 2008; Headey, Schupp, Tucci and Wagner, 2010). Similarly, it has been reported that individuals who prioritize altruistic and family goals are on average happier than those who prioritize financial and

material goals (Nickerson, Schwarz, Diener and Kahneman, 2003; Headey, 2008).

In previous papers, replicating results with Australian, British and German panel data, we found that LS is significantly affected by the personality of one's partner, and by behavioural choices relating to social networks and participation in social activities, frequency of physical exercise, and achieving a preferred balance between work and leisure (Headey, Muffels and Wagner, 2010, 2013). The links between these behavioural choices and LS held in within-person (fixed effects) panel regression models, confirming that changes in choices covary over time with changes in LS.

A weakness of previous efforts to account for change in LS (including our own) is that, while they may provide explanations for short term change, they do not directly address the issue of how and why LS may remain substantially above (or below) its previous level in the medium or long term, rather than reverting to its previous level, as set-point theory would predict.

A causal model: Granger-causation with positive feedback loops

We attempt here to account for medium term change with two-way causation models (structural equation models), which include positive feedback loops between LS and variables usually treated as causes of LS. It must be conceded that disentangling two-way causation remains problematic, even with long term panel data. Previous researchers have claimed to find two-way causation between LS and domain satisfactions (mainly marriage and job satisfaction), and between LS and quality of social networks, health, housing and volunteering (Headey, Veenhoven and Wearing, 1991; Thoits and Hewitt, 2001; Meier and Stutzer, 2004; Mathison et al, 2007; Nagazato, Schimmack and Oishi, 2011). These studies have been criticised by researchers who claim that structural equation models involving two-way causation can be unstable in the sense that small, reasonably plausible changes in model specification can lead to quite different conclusions about direction of causation (Scherpenzeel and Saris, 1996; see also Wooldridge, 2010).

In this paper we suggest that more stable and better fitting models of LS can be estimated by conceptualising relationships in terms of *Granger-causation* (Granger, 1969; Granger and Newbold, 1974; Pearl, 2009).

The intuitions behind Nobel Laureate, Clive Granger's concept of causation, are that (1) causes must precede effects and (2) if earlier values of an x variable (a presumed cause) can be shown to be statistically

significantly related to later values of a y variable (a presumed effect) in equations which include multiple lags of y , then it may be inferred that x is one cause of y . Granger's view is that, by including multiple lags of y , the researcher usually takes care of most of the variance due to omitted variables ('unobserved heterogeneity'), and so avoids biased estimates and statistically significant autocorrelated error terms that indicate longitudinal bias (Granger, 1969; Pearl, 2009; Wooldridge, 2010; Beck and Katz, 2011; Wilkins, 2014).

We suggest extending Granger's approach to two-way causation and panel data, and being willing to infer that x and y may cause each other in models of the kind depicted in Figure 1. This model shows possible relationships between health (the x variable) and LS (the y variable). It is based on five waves of panel data, in line with the models to be estimated in this paper.

INSERT FIGURE 1 HERE

The causal paths of main interest in Figure 1 – usually referred to in the causal modelling literature as *cross-lagged paths* - are those with BU? and TD? above them. In referring to bottom-up (BU) and top-down (TD) paths, we use terminology coined by Diener (Diener, 1984; Diener et al,

1999). He labels models in which LS is viewed as being caused by a weighted combination of x variables as BU models, and models in which evaluations of specific aspects of life are viewed as consequences (or spin-offs) of overall LS as TD models.

The curved arrows marked SP? in Figure 1 are a further attempt (in addition to Granger lags) to allow for the possibility that covariance between health and LS could be wholly or partly spurious. These curved arrows link the error terms (residuals) of each pair of x and y equations, and should be routinely included in structural equation panel models (Kessler and Greenberg, 1981; Finkel, 1995; Wooldridge, 2010).

‘Extra’ Granger lags are shown by dotted arrows running along the top and bottom of Figure 1. As well as helping to avoid biased estimates of the coefficients of main interest, these extra lags can be given a substantive interpretation in LS research. If ‘extra’ lags of LS are significant in LS equations, it means that individual happiness at time t depends not just on happiness at $t-1$, but is also affected in a cumulative way by happiness in earlier periods. Evidence about the effects of lagged x variables on x at time t can be given a similar substantive interpretation.

In summary, four main outcomes are possible when we estimate Granger-style models as shown in Figure 1. One possibility is that the covariance

between a particular x variable and LS will be found to be wholly spurious. Secondly, it is possible that only BU paths from x to LS will prove to be statistically significant. Alternatively, only TD paths from LS to x may be significant. If both BU and TD paths prove to be significant, then the inference will be that changes in x and changes in LS cause and reinforce each other in positive feedback loops.

METHODS

This section is in three parts. We briefly describe the three national household panel surveys. Then we describe the main measures: LS and the x variables. Then comes a longer section on issues involved in modelling two-causation.

Three long-running national household panel surveys

The Household Income and Labour Dynamics Australia (HILDA) Panel Survey

The HILDA panel began in 2001 with a sample of 13,969 individuals in 7,700 households (Watson and Wooden, 2004). Interviews were achieved in 61% of in-scope households. All household members age 15 and over are interviewed. The cross-sectional representativeness of the panel is maintained by interviewing ‘split-offs’ and their new families. So when a young person leaves home (‘splits off’) to marry and set up a new family,

the entire new family becomes part of the panel. It may be noted that, as happens in many panels with good retention rates, the sample size is now increasing. That is, the number of individuals added to the panel each year, via split-offs and young people turning 15, exceeds the number who die, cannot be traced, or drop out by refusing an interview. In this paper we use annual data for 2001-12.

The British Household Panel Survey (BHPS)

The British (BHPS) panel was launched in 1991 with about 10,300 individuals in 5,500 households (Lynn, 2006). However, a question about life satisfaction was not included until 1996, so in this paper only 1996-2008 data are used. All individuals in sample households who are age 16 and over are interviewed. Again, sample representativeness is maintained by including split-offs and their new households. The British panel was augmented by booster samples for Scotland and Wales in 1991 and a new Northern Ireland sample in 2001. In 2008, the latest year used in this paper, the sample size was just over 14,000.

The German Socio-Economic Panel Survey (SOEP)

SOEP was the first of these panels to get underway. It began in 1984 in West Germany with a sample of 12,541 respondents (Frick, Schupp and Wagner, 2007). Interviews have been conducted annually ever since.

Everyone in the household age 16 and over is interviewed. Split-offs and

their new family members (if any) join the panel. The sample was extended to East Germany in 1990, shortly after the Berlin Wall came down, and since then has been boosted by the addition of new immigrant samples, a special sample of the rich, and recruitment of new respondents partly to increase numbers in ‘policy groups’ (e.g. welfare recipients). There are now over 60,000 respondents on file, including some grandchildren as well as children of the original respondents. The main topics covered in the annual questionnaire are family, income and labour force dynamics. Questions on LS, domain satisfactions (job satisfaction, satisfaction with income etc), health, social participation and exercise have been included every year. Data for 1984-2010 are used in this paper.

Measures

Life Satisfaction (LS)

LS is measured on a 0-10 (‘totally dissatisfied’ to ‘totally satisfied’) scale in HILDA (mean=7.910 sd=1.510) and also in SOEP (mean=7.001 sd=1.834). In the BHPS a 1-7 scale is used (mean=5.223 sd=1.299).

Single item measures of LS are plainly not as reliable or valid as multi-item measures, but are internationally widely used in household panel surveys and have been reviewed as acceptably reliable and valid (Diener et al, 1999).

Variables potentially implicated in two-way causation with LS

Specific measures of the x variables in this paper differ in non-trivial ways among the three surveys. Because researchers in Melbourne, Colchester and Berlin happen to have selected different measures, we have a particularly stringent test of whether our two-way causation results replicate cross-nationally.

Health

The main health measures in HILDA are derived from the internationally widely used SF-36 Health Outcomes Survey (Ware, Snow and Kosinski, 2000). The 36 items cover physical and mental health issues. The data are factor analysed to yield physical and mental health summary scales. The physical health scale used here is mainly based on questions about whether, “Your health now limits you in these activities ...lifting or carrying groceries...climbing one flight of stairs...bathing or dressing yourself?” Also included are self-assessment items (e.g. “In general, would you describe your health as excellent, very good, good, fair or poor?”). Scores are standardized to range from 0 to 100. The scale’s correlation with LS in HILDA is 0.322.

The BHPS and SOEP have only included short versions of the Health Outcomes Survey in recent years. However, they have for many years included a very simple, but internationally widely used, self-rated health measure on which respondents rate their own health on a 1-5 scale ('very poor' to 'excellent'). Despite its simplicity, or perhaps because of it, this scale has been assessed as reasonably valid in that it correlates satisfactorily with physician ratings (Schwarze, Andersen and Silke, 2000). Its correlation with LS in the BHPS is 0.279 and in SOEP it is 0.378.

Social support and participation

Social network measures are intended to assess the quality of social support available to a person. The social support measure included in HILDA defines support in terms of availability of intimate attachments, friendships and alliances (Henderson, Byrne and Duncan-Jones, 1981). The ten survey items, which are asked on a 1-7 scale, can be combined into a single measure of quality of *social support*. Typical items are: "When I need someone to help me out, I can usually find someone" and "I don't have anyone I can confide in". The correlation of social support with LS is 0.371.

The BHPS and SOEP only measure social networks intermittently, rather than annually. However, they include annual measures of social participation and, in particular, measures of interaction with friends and neighbours. These are measures of flow/interaction, whereas HILDA measures a stock; that is, the quality of a person's support network.

For the British panel, our social participation index combines two correlated items relating to frequency of 'meeting with friends and relatives' and frequency of 'talking with neighbours'. These are asked on a response scale running from 'on many days' (code 1) to 'never' (code 5). The correlation of this index with LS is 0.092. The social participation index used for Germany combines two correlated items about frequency of 'meeting with friends, relatives or neighbours' and 'helping out friends, relatives or neighbours'. The response scale has three points: 'every week', 'every month' and 'seldom or never'. The correlation of this index with LS is 0.133.

Exercise

Frequency of exercise is usually found to be associated with LS, as well as longevity (Gremeaux et al, 2012). HILDA respondents are asked annually about how frequently they engage in moderate or intensive physical activity lasting for at least 30 minutes (ABS, 1996). The

response scale runs from 0 ('not at all') to 5 ('every day'). The correlation of this item with LS is 0.109. SOEP respondents are asked a quite different question about how frequently they engage in active sport or exercise. The response scale runs from 0 ('not at all') to 5 ('every day'). This item's correlation with LS is 0.185. In the British panel the comparable question is about how often respondents walk, swim or play sport. The 5-point response scale runs from 'at least once a week' to 'never/almost never'. Its correlation with LS is 0.093.

Job satisfaction, satisfaction with one's partner and satisfaction with family life

Satisfaction with various domains of life is measured annually in all three surveys. Mindful of Freud's dictum that, "To be happy is to work and to love", we include job satisfaction and satisfaction with one's partner or family life in our two-way analyses. It is surely reasonable to hypothesise (perhaps not in line with Freud) that satisfaction in these domains could be caused by, as well as cause LS.

HILDA and SOEP measure job satisfaction with a single question asked on the same 0-10 ('totally dissatisfied' to 'totally satisfied') scale as LS, while the BHPS uses its 1-7 scale (as for LS). The correlations of these

items with LS are 0.419 (Australia), 0.473 (Britain) and 0.402 (Germany).

HILDA and the BHPS include single item measures of satisfaction with one's spouse or partner (again using their different scales). These items correlate quite strongly with LS: 0.379 (Australia) and 0.253 (Britain). In SOEP a measure of partner satisfaction has only been included in the last few years, so we instead use a longer-running measure of satisfaction with family life (0-10 scale), which has a correlation with LS of 0.462.

Exogenous variables included as 'controls' and also to assist with model identification

Additional exogenous variables are included in the five-wave panel models, both as standard 'controls' and also to assist with model identification. Standard 'controls' are: gender (female=1 male=0), age, age-squared, age-cubed, partner status (partnered=1 not partnered=0), level of education, household net income, unemployed (unemployed=1 other=0), disability status (disability=1 other=0).¹ It is also desirable to control for personality traits known to be correlated with LS (Lucas, 2008). Traits measured in all three panels are the so-called Big Five, which many psychologists regard as adequately describing normal or

¹ Disability status was not included in models in which the health scale was the x variable.

non-psychotic personality: neuroticism, extroversion, openness to experience, agreeableness and conscientiousness (Costa and McCrae, 1991). Since the traits are partly genetic (Lucas, 2008), it clearly makes sense to treat them as exogenous and causally antecedent to LS and the x variables in our models.

In any panel survey, what are called ‘panel conditioning effects’ are a possible source of bias. That is, panel members might tend to change their answers over time – and answer differently from the way non-panel members would answer - as a consequence just of being in a panel. There is evidence for all three panels that panel members, in their first few years of responding, tend to report higher LS scores than when they have been in the panel for a good many years (Frijters, Haisken-DeNew and Shields, 2004; Headey, Muffels and Wagner, 2013). This could be due to ‘social desirability bias’; a desire to look good and appear to be a happy person, which is stronger in the first few years of responding than in later years. Or it could be due to a ‘learning effect’; learning to use the middle points of the 0-10 scale, rather than the extremes and particularly the top end.

To compensate for these possible sources of bias, we include in all equations a variable which measures the number of years in which each panel member has already responded to survey questions.

Data analysis: structural equation modelling of two-way causation

The structural equation models estimated here are based on five-year overlapping periods. So for example, in the German panel, the five-year periods are 1991-95, 1992-96...2006-10. The reason for using overlapping periods, instead of separate five-year blocks, is to obtain more reliable results due to larger sample numbers. An assumption which has to be met for this decision to be sensible is that relationships among variables do not change much within the overall time period. Inspection of bivariate correlations within and across waves suggests that this assumption is plausible. A further point is that sample numbers in all analyses relate to *person-years*, rather than persons.

Structural equation modelling, rather than OLS regression analysis, is required whenever the aim is to estimate a set of equations, rather than a single equation, and especially when two-way causal links are involved.²

The structural equations in this article are estimated using maximum likelihood analysis.³ Maximum likelihood coefficients and standard errors can be given the same interpretation as regression coefficients. However, assessing the ‘goodness of fit’ of structural models is more complicated

² Regression analysis is essentially a single equation technique. Regression estimates derived from multi-equation systems are likely to be biased, due to correlations between explanatory variables and error terms in some or all equations. A key assumption of OLS regression is that such correlations are zero.

³ ML estimates are usually consistent and asymptotically normal under the (not very restrictive) assumption of *conditional normality* (StataCorp, 2013).

than for regression models. Several measures of fit are conventionally used. The root mean squared error of approximation (RMSEA) and the standardized root mean residual (SRMR) are directly based on comparing differences (residuals) between the actual input matrix with the matrix implied by model estimates. It has become conventional to regard an RMSEA under 0.05 and an SRMR under 0.08 as satisfactory (Bentler, 1990; Browne and Cudeck, 1993).

More complicated assessments are provided by the Comparative Fit Index (CFI) and the Tucker-Lewis index (TLI). The CFI is based on a likelihood ratio (LR) chi-square test and takes account of the contribution of each estimate in the model to overall goodness of fit. The TLI is also derived from an LR chi-square test, and is particularly useful because it rewards parsimony and penalises models including explanatory variables which account for little variance, even if statistically significant. CFI and TLI fits above 0.90 used to be regarded as satisfactory, but some recent reviews recommend 0.95 (Bentler, 1990; Browne and Cudeck, 1993).

Models involving two-way causation can be unstable in the sense that they would never reach equilibrium and would 'blow up' if sufficient

iterations were run (Finkel, 1995; StataCorp, 2013).⁴ In view of this, Bentler and Freeman (1983) developed a test of model stability, which is used here.

Strictly speaking, maximum likelihood estimation of structural equations requires an assumption that the endogenous variables are measured on an interval or ratio scale. In fact, all the endogenous variables in our equations (LS and the x variables) are measured on fairly long ordinal scales. However, it has become routine in research on LS to treat the data as interval-level. Andrews and Withey (1976) were the first researchers to show that, substantively, results using interval-level statistics were much the same as those using ordinal statistics. Most researchers since have followed their lead. An important practical reason for making interval scale assumptions in structural equation modelling is that, although equations can be estimated for models with ordinal or binary endogenous variables, few measures of model fit are available, so it is often unclear whether one model is statistically preferable to another.⁵

Model identification

⁴ Model ‘stability’ is here used as a technical term. We previously used the term in a different sense to refer to Scherpenzeel and Saris’s (1996) claim that two-way causation models of LS are unstable because apparently small differences in model specification can lead to substantially different estimates.

⁵ Another limitation is that covariances between the error terms of equations cannot be estimated, so it becomes difficult to assess whether relationships are spurious.

Special steps are always needed to achieve model identification in estimating two-way causation. There are three main approaches, all of which are implemented in this paper:-

(1) Exogenous variables may serve as instrumental variables. In our models some of the exogenous variables described earlier – specifically, gender and personality traits – are fixed characteristics (i.e. fixed intra-individually) which we link to wave 1 measures of x and LS, but not to later waves. The rationale for omitting links to later waves is that we would not expect fixed characteristics to influence later measures, net of their effects on wave 1 variables (Kessler and Greenberg, 1981).

Equivalently, we may say that there is no reason to expect fixed exogenous characteristics to be associated with *changes* over time in x and LS.

(2) Lagged versions of x and LS have served as instruments in previous research (Scherpenzeel and Saris, 1996; Kuskova, 2011; Nagazato, Schimmack and Oishi, 2011). However, in our models the inclusion of multiple Granger-style lags increases rather than reduces the number of free parameters to be estimated. Our models are nevertheless still identified, due to inclusion of exogenous variables. Further, in later computer runs we modified our models by removing some Granger-style

lags – specifically longer term lagged effects of x on LS, and LS on x .

Consequently, in our final models, this second approach to model identification came back into play.

(3) Equality constraints may be imposed. That is, sets of coefficients may be fixed equal to each other, so reducing the number of free parameters to be estimated (Kessler and Greenberg, 1981; Finkel, 1995). No equality constraints were imposed in our initial models. However, both a priori reasoning and initial computer runs indicated that, empirically, some causal links appeared to be almost exactly the same in consecutive waves of data. So equality constraints were added in later runs, and in some cases improved model fit.

Because we make use of all three approaches to achieving identification in panel data models with two-way causation, our models are ‘over-identified’; that is, they actually contain many more bits of information than are required to estimate parameters. This is ideal for maximum likelihood estimation; the maximum likelihood estimator is designed to find the best solution among the range of solutions available.

RESULTS

Trajectories of LS

A revised theory needs to be able to account for typical trajectories of LS over time.

Here are the trajectories of three German respondents (for illustrative purposes it is convenient to use the longest-running of the national panels), selected because their actual trajectories in 1991-2010 are typical of individuals in the top, middle and bottom thirds of the LS distribution. Of course, no three individuals can perfectly represent a large sample, but these three are typical in having 20-year means and standard deviations of LS which are exactly average for ‘their’ third of the distribution.⁶

INSERT FIGURES 2.1-2.3 HERE

Set-point theory predicts that individuals will have a *flat, long-term trajectory of LS*, except for temporary fluctuation due to life events. If an upward or downward fluctuation occurs at time t , the theory predicts that the individual will revert to set-point at $t+1$. The most obvious point to

⁶ It is accepted, of course, that differing long-term trajectories can have the same mean and standard deviation. Our inspection of a very large number of trajectories indicated that few panel members recorded continuous long-term declines or continuous long-term increases in LS. The majority recorded trajectories with multi-year periods of both relatively high and relatively low LS, as shown in Figures 2.1-2.3. The statistical analyses below confirm this point.

make about the trajectories of the typical individuals portrayed in Figures 2.1-2.3 is that they are volatile, and it is far from clear that they have any set-point. Of course, one could calculate their individual mean (or median) ratings for the 20-year period, but the evidence does not suggest that they have fixed set-points to which they keep reverting.

We graphed the LS trajectories of many individuals in the three national panels. The evidence led us to hypothesise that, instead of reverting to a set-point, many people spend several consecutive years above... and then several years below their own long term mean. This hypothesis was tested by a series of within-person (fixed effects) panel regression analyses (see Table 1). In these analyses the independent variable was LS in year t . The dependent variables were successively (i) LS in years $t+1$ and $t-1$ (ii) LS in years $t+2$ and $t-2$, and (iii) LS in years $t+3$ and $t-3$. In order to interpret the results, it is crucial to appreciate that, in these fixed effects analyses, coefficients are calculated by relating annual deviations from each individual's own long-term grand mean on the dependent variable to annual deviations from his/her grand mean(s) on independent variable(s).

INSERT TABLE 1 HERE

The coefficients in Table 1 indicate that, in the long-running German panel, individuals who are above their own long term mean of LS in year t are more rather than less likely to be above it for each of the three years beforehand, and the three years afterwards; six years in total.⁷ Similarly, individuals below their 20-year mean of LS in year t are likely to remain below it for six years. These results are clearly counter to set-point theory, which as we just noted, predicts negative relationships between time t and $t+1$ scores for individuals who have deviated from their own long term mean at time t .

In the shorter periods covered by the Australian and British panels, it appears that respondents typically spend just four consecutive years, either above or below their own long term mean. However, it is quite likely that, when more years of data become available, the Australian and British results will be close to the German. Clearly, the longer the time period, the greater the probability of individuals remaining in the same 'state' for multiple consecutive years. It may be noted that, if analysis of the German data is restricted to 12-13 years (the periods covered by the Australian and British panels), results indicate just four (not six) year periods above or below the mean.

⁷ However, given the relatively small size of the coefficients, there are certain to be many individuals who do not remain above (or below) their own mean for *all* of the six years.

These fixed effects regression results have quite important implications for a theory of LS. They show that LS tends to change in medium term oscillations, rather than fluctuating short term around a stable set-point.

Structural equation models with 2-way causation: positive feedback loops between LS and x variables

In the rest of the paper we attempt preliminary explanations of why these medium term changes occur. We indicate that there appears to be two-way causation between LS and a range of x variables, so that positive feedback loops sometimes tend to perpetuate periods of above average LS, and sometimes perpetuate below average periods.

In analysing the five-wave panel data (see Figure 1 above), we first estimated relationships between LS and each particular x variable, deploying a Granger-style model with multiple lags. We then modified this model, removing statistically insignificant links and imposing equality constraints where appropriate. Our final Granger-style model - both substantive causal estimates and measures of model fit - was compared with results from one-way causal models, and also from cross-lagged and simultaneous causation models of the kind estimated in previous studies which have investigated two-way causation.

Health and LS

Our first step was to estimate a ‘full’ Granger-style model; that is, a model in which all lags of both health scales and of LS were included in equations. It immediately became clear that, in all three national datasets, all lags of the health measure were statistically significant in health equations, and all lags of LS were significant in LS equations. This already suggested that the Granger approach was going to improve model fit, compared with previous approaches. It was also clear, however, that some paths could be eliminated from the model. While health lagged by one time period had a statistically significant effect on LS, and LS lagged by one period had a significant effect on health, little additional variance was accounted for in any dataset by ‘extra’ (2nd, 3rd etc) cross-lags. No extra cross-lags of LS were significant in any of the three national datasets, and although some two-year cross-lags of self-rated health were significant at the 0.05 level, their effects were substantively trivial. So we dropped extra cross-lags from final models.

We also imposed equality constraints on the one-year cross-lagged links of main interest, marked α and β in Figure 1 (above).⁸ The rationale

⁸ In the final run of these models, the equality constraints on the BU and TD estimates for the equations for Health₂ and LS₂ were dropped. The reason is that these are not ‘Granger’ equations in that no ‘extra’ (2nd, 3rd, etc) lags are available. Consequently, as Granger would predict, the estimates of the

was that inspection of correlation matrices for each country indicated that relationships between self-rated health and LS were similar within each wave of data, and also between consecutive waves. So it seemed reasonable to hypothesise that ‘true’ relationships were, in fact, constant or almost constant over time. In the event, for all three countries, a model with equality constraints was neither clearly a better, nor clearly a worse fit than a model without these constraints. It depended on which measures of fit one chose to place faith in. The LR Chi-square test and the CFI diagnosed a somewhat worse fit with the equality constraints in place, but the RMSEA and TLI tests, which reward model parsimony, diagnosed an improved fit. On grounds of parsimony, we elected to treat the Figure 1 model with equality constraints as our preferred model.

Table 2 for each country gives both metric and standardized maximum likelihood estimates for the relationships of main interest (BU, TD and SP) in Figure 1. The standardized results are particularly useful because they enable us to compare effect sizes. The tables also provide a fairly comprehensive set of measures of model fit.

Insert Table 2 for each country here

BU and TD links from these equations are considerably higher than from the equations with multiple lags, and are probably biased (Granger and Newbold, 1974).

Two-way causal links are found between self-rated health and LS. A comparison of the standardized coefficients indicates that, in the Australian and German data, the BU links (=a in Figure 1) from Health to LS are considerably stronger than the TD links (=b in Figure 1) from LS to Health. The Australian two-way links are BU=0.073 ($p<0.001$), compared with TD=0.018 ($p<0.001$). The equivalent German results are BU=0.079 ($p<0.001$), compared with TD=0.057 ($p<0.001$). In Britain the BU and TD effects appear about equally strong (BU=0.062 $p<0.001$, TD=0.069 $p<0.001$).

More detailed analyses (not printed here) showed that the relative effect sizes of BU and TD links in all three countries are approximately the same for men and women, and for older and younger people.⁹ Not unexpectedly, SP links, reflecting the effects of omitted variables, are substantial in all three datasets.¹⁰

These Granger-style models have satisfactory fits to the data. The models for each country have CFIs and TLIs which are comfortably above the standard ‘close fit’ criterion of 0.95. The RMSEAs are all satisfactorily

⁹ Results for sub-sets of the population are not printed here; available from the authors.

¹⁰ There are five of these correlated error terms: one for each wave of data. The tables report correlations for the fifth wave. Among the omitted variables which may account for the high SP coefficients are the other x variables included in later analyses. We partially deal with this issue in multivariate models estimated later in the paper.

below the standard cut-off of 0.05, and the SRMRs are well below the cut-off of 0.08. No significant autocorrelated error terms are found. Model stability is satisfactory; all eigenvalues are within the unit circle (Bentler and Freeman, 1983).

Additional sensitivity tests were performed to check whether the equality constraints in the model are justified. The results were slightly ambiguous. It transpired that, if all constraints were removed, estimates of the BU and TD paths of main interest all remained within 0.02 (standardized) of the estimates reported in these tables. Lagrange multiplier tests indicated that, in each country, model fit would be slightly improved by removing just one pair of imposed equalities. For example, in the model for the BHPS data, the removal of the imposed equality between (i) the path from LS in Year 2 to Health in Year 3 and (ii) the path from LS in Year 4 to Health in Year 5...would result in re-estimated paths which were significantly different from each other, although only at the 0.01 level. On grounds of theory and parsimony, we elected not to make this change and to retain the model shown in Figure 1.

It is crucial to compare our preferred model with all reasonable alternatives, including one-way causal models and also cross-lagged and simultaneous causation models of the kind deployed in previous research.

Here it should be mentioned that all alternative models are ‘nested’ versions of a larger Granger-style model, so model fit can be directly compared (Bentler, 1990). It transpired that the Granger models are a much closer fit to the input data than alternatives. In all three countries a model with only one-way causation from health to LS (still including ‘extra’ Granger lags and equality constraints) is a much worse fit to the data. For example, in the German case, a one-way model yields exactly the same standardized estimate of the BU link from health to LS as the Granger model, but is a much worse fit to the data: CFI=0.897, TLI=0.862, RMSEA=0.061 and SRMR=0.039. A one-way TD model with causation running only from LS to health also gives the same estimate of substantive interest as the Granger model, but is an even worse fit: CFI=0.892, TLI=0.854, RMSEA=0.068 and SRMR=0.049.

The Granger model is also a closer fit to the data than either of the two-way causation models commonly estimated in previous LS research, namely cross-lagged models without ‘extra’ Granger lags and simultaneous causation models without extra lags. In the German case, a cross-lagged model gives the following fit readings: CFI=0.895, TLI=0.859, RMSEA=0.067 and SRMR=0.060. Readings for a simultaneous causation model are: CFI=0.858, TLI=0.812, RMSEA=0.077 and SRMR=0.071.

Despite fitting the data less well, it should be pointed out that, in all three countries, estimates of both BU and TD relationships in the cross-lagged and simultaneous causation models are statistically significant, indicating two-way causation. The estimates are, in fact, somewhat higher than in our preferred model, but that is a consequence of omitting extra ‘Granger’ lags; an omission which clearly worsens model fit.

Social support, social participation and LS

Full multiple-lag Granger models for links between LS and social support (Australia) and social participation (Britain and Germany) indicated that both first and second lags of BU and TD relationships were statistically significant. However, the effect sizes of the second lags were small, so they were omitted from final models. Again, as with the health models, inspection of the input correlation matrices made it clear that relationships within and between waves of data were quite similar over time. So we again estimated models in which, as in Figure 1, the cross-lagged causal links of main interest were constrained to be equal.

Insert Table 3 for each country here

The results indicate small, but clearly statistically significant two-way links ($p < 0.001$) in all three countries. However, effects appear stronger in the Australian data (BU=0.085 $p < 0.001$, TD=0.056 $p < 0.001$) than in the other two countries. This is probably because the measure taken in Australia is a social network stock measure, rather than a measure of current social participation.¹¹ The SP link is again significant in all three countries ($p < 0.001$).

Measures of fit again do not give an unambiguous message as between the full Granger-style models and the reduced models. Measures which reward parsimony indicate that the reduced models are marginally preferable, whereas measures which give less weight to parsimony suggest that the full models are a slightly better fit. BU and TD estimates for both models in all three countries are approximately the same, except that in the full Granger models the BU and TD variance accounted for gets partitioned between first and second lags of the explanatory variables, rather than being entirely attributed to first lags.¹² As was the case with the health models, Lagrange multiplier tests diagnosed just one

¹¹ Again, no statistically significant differences in these relationships were found between men and women, or between older and younger people.

¹² For example, the German data yield the following estimates: in the full Granger model the first BU lag (standardized) = 0.059 ($p < 0.001$) and the second BU lag = 0.019 ($p < 0.001$). The first TD lag = 0.045 ($p < 0.001$) and the second lag = 0.010 ($p < 0.05$).

pair of imposed equalities in each dataset as unjustified; again, on grounds of theory we chose to retain the Figure 1 model.

For all three countries, alternative one-way causation models are a much worse fit to the input data. For example, for Australia a BU model with links only from social networks to LS has these fit readings: CFI=0.783, TLI=0.665, RMSEA=0.100 and SRMR=0.068. A one-way TD model has these readings: CFI=0.848, TLI=0.767, RMSEA=0.095 and SRMR=0.071. Cross-lagged and simultaneous models without ‘extra’ Granger lags yield two-way causal estimates which are larger (in the case of the simultaneous models much larger) than the Granger-style models, but the alternative models are poor fits to the data. For example, for Australia the cross-lagged and simultaneous models both have CFIs of 0.837, the TLIs are 0.787 and 0.788 respectively, and the RMSEAs and SRMRs are unsatisfactory and much higher than for the Granger-style models. Furthermore, the SP links in the simultaneous models are substantial and negative (approximately -0.200). Negative correlated errors are almost certainly a symptom of model misspecification (Finkel, 1995). It is hard to conceive of any omitted variables which could somehow generate negative links between LS and either social support or social participation.

Exercise and LS

Evidence from one-way causation models usually indicates that exercise has a modest but statistically significant positive effect on LS (Argyle, 2001; Headey, Muffels and Wagner, 2010, 2012). Table 4 for each country gives results for two-way models.

Insert Table 4 for each country here

In all three countries the effect of exercise on LS is considerably stronger than the reverse effect. The standardized estimates for the effects of exercise are 0.031 (Australia), 0.035 (Britain) and 0.034 (Germany); all significant at the 0.001 level. The measures of fit in the right hand columns of each table diagnose very close fits to the data. Again, no statistically significant autocorrelated errors were found. Just one pair of imposed equalities for each country is found not to be strictly justified, but the overall model, assessed by a likelihood ratio (LR) test, would only be slightly improved if these constraints were lifted.

Alternative one-way causation models, and also two-way cross-lagged and simultaneous causation models, yield BU and TD estimates which

are larger but broadly consistent with estimates for our preferred model. Again, however, the fit of these alternative models makes them unacceptable.

A multivariate two-way causal model: links between health, social support, social participation, exercise and LS

To this point we have analysed two-way causation between LS and x variables, taking one x at a time. However, it is a reasonable hypothesis that health, social support (or social participation) and exercise exert more or less simultaneous, combined effects on LS, and so should be entered at the same step in a causal model. Table 5 for each country gives results for our preferred multivariate model with imposed equality constraints.

Insert Table 5 for each country here

It transpires that in these multivariate models health, social support (or participation) and exercise all still have statistically significant ($p < 0.001$) BU and TD links with LS. It is still the case that the standardized BU paths between health and LS are larger than the TD paths in Australia and Germany, but not in Britain. Both social support/participation and

exercise are still shown to have modest but statistically significant two-way links with LS.

It is often the case that large structural equation models are a poor fit to the data. It is pleasing that, for all three countries, these relatively large models have satisfactory fits.¹³ Only the TLI result for Australia (0.948) drops marginally below what is regarded as the cut-off point for a satisfactorily close fit (0.950).

LS related to job satisfaction and satisfaction with partner or family life

In a final model we estimate the combined effects of job satisfaction and satisfaction with one's partner or family life on LS. Estimates given in Table 6 are from our preferred model with constrained equalities.

Insert Table 6 for each country here

Significant two-way causation is found between both these domain satisfactions and LS. However, the standardized TD links between LS

¹³ As was the case for models with only one x variable, some imposed equality constraints in these multivariate models (in fact, 2 out of 18 in each dataset) were diagnosed as not strictly justified. Again, however, the measures of fit which reward parsimony – the TLI and RMSEA – provided countervailing evidence in favour of retaining the constraints. On grounds of theory we preferred to keep them.

and job satisfaction are considerably stronger in all three countries than the BU links, although the BU links are still significant at the 0.001 level. This suggests that happy people tend to feel happy with their jobs, rather than that job satisfaction makes a big contribution to LS (see also Headey, Veenhoven and Wearing, 1991; Lance, Mallard and Michalos, 1995). Results for the links between LS and satisfaction with one's partner or family life differ slightly among these countries. In the Australian dataset the BU link from partner satisfaction to LS is somewhat stronger than the reverse link, whereas in Britain and Germany it is the other way round.

These domain satisfaction models all have good fits to the input data: CFIs and TLIs are all above 0.95, RMSEAs are below 0.05, and SRMRs are below 0.08.

DISCUSSION

The two-way causal results - the BU and TD effects - reported in this paper replicate quite strongly in Australia, Britain and Germany.

Replication occurs despite non-trivial differences in question wording and the response scales used in the three countries; an indication that our results are quite robust. It appears that LS is linked in two-way relationships with quite a wide range of variables: behavioural choices

(frequency of exercise and active social participation), domain satisfactions (job satisfaction and marital satisfaction), social support and health. It seems reasonable to suggest that, in the LS field, two-way causation may be pervasive and should be regularly investigated in future research.

Adapted Granger-style models have provided a plausible and consistent account of the two-way relationships investigated here. The models have a satisfactory fit to the data, and in fact are a much closer fit than alternative one-way causation, cross-lagged and simultaneous causation models of the kind estimated in previous research. It is nevertheless worth recording that these alternative models, despite being a worse fit to the data, also indicate two-way causation.

Many researchers in the LS field now believe that our main current challenge is to develop a theory of change. Set-point theory is purely a theory of stability and fails to account for evidence from panel surveys, showing that the LS trajectories of many respondents are subject to medium and/or long term change.

It seems quite likely that analysis of two-way causation - *positive feedback loops* – can make a substantial contribution towards developing

a theory of change. If our models are valid, they provide a partial explanation of why, instead of have a stable set-point, the LS trajectories of many individuals show that they go through ‘happy times’ and ‘unhappy times’. That is, they spend extended periods above their own long term average LS, and other extended periods below their average.

Plainly, we have only made a few steps towards accounting for change. Our analyses are based on examining graphs of medium and long term trajectories of LS. Clearly, many trajectories are possible. In future work, it will be important to classify trajectories and attempt to explain the patterns of change that they reveal.

References

Andrews, F.M. and Withey, S.B. (1976) *Social Indicators of Well-Being*. New York, Plenum.

Argyle, M. (2001) *The Psychology of Happiness*. London, Routledge.

Australian Bureau of Statistics (ABS)(1996) *National Health and Nutrition Survey, 1995*. Canberra, ABS.

Beck, N. and Katz, J.N. (2011) Modelling dynamics in time-series cross-section political economy data, *Annual Review of Political Science*, 14, 331-52.

Bentler, P.M. (1990) Comparative fit indices in structural models, *Psychological Bulletin*, 107, 238-46.

Bentler, P.M. and Freeman, E.H. (1983) Tests for stability in linear structural equation systems, *Psychometrika*, 143-45.

Brickman, P.D. and Campbell, D.T. (1971) 'Hedonic relativism and planning the

good society' in M.H. Appley ed. *Adaptation Level Theory*. New York,
Academic
Press, pp. 287-302.

Browne, M.W. and Cudeck, R. (1993) 'Alternative ways of assessing
model fit'
in K.A. Bollen and J.S. Long eds. *Testing Structural Equation Models*.
Newbury Park, Ca., Sage, pp.136-62.

Costa, P.T. and McCrae, R.R. (1991) *NEO PI-R*. PAR, Odessa, Fla.

Deeg, D. and van Zonneveld (1989) 'Does happiness lengthen life?' In
R.Veenhoven ed. *How Harmful Is Happiness?* Rotterdam, Erasmus
University Press, chap. 5.

Diener, E. (1984) Subjective Well-Being, *Psychological Bulletin*, 95,
542-75.

Diener, E. and Seligman, M.E.P. (2004) Beyond money: Toward an
economy of
well-being, *Psychological Science in the Public Interest*, 5, 1-31.

Diener, E., Suh, E.M., Lucas, R.E. and Smith, H.L. (1999). Subjective well-being: Three decades of progress, *Psychological Bulletin*, 25, 276-302.

Dunn, E.W., Aknin, L.B. and Norton, M.I. (2008) Spending money on others promotes happiness, *Science*, 319, 1687-88.

Easterlin, R.A. (2003) Explaining happiness, *Proceedings of the National Academy of Sciences*, 100.19 (May 23) 11176-11183. doi: 10.1073/pnas.1633144100.

Finkel, S.E. (1995) *Causal Analysis With Panel Data*. Thousand Oaks, Ca., Sage.

Frey, B.S. and Stutzer, A. (2002) What can economists learn from happiness research? *Journal of Economic Literature*, 40, 402-35.

Frick, J.R., Schupp, J. and Wagner, G.G. (2007) Enhancing the power of the German Socio-Economic Panel Study (SOEP) – evolution, scope and enhancements, *Schmollers Jahrbuch*, 127, 139-69.

Frijters, P., Haisken-DeNew, J.P., and Shields, M.A. Money does matter! Evidence from increasing real income and life satisfaction in East

Germany following reunification, *American Economic Review*, 94, 730-40.

Fujita, F. and Diener, E. (2005) Life satisfaction set-point: Stability and change.

Journal of Personality and Social Psychology, 88, 158-64.

Granger, C. W. J. (1969) Investigating causal relations by econometric models and cross-spectral methods, *Econometrica*, 37.3, 424–38.

Granger, C. W. J. and Newbold, P. (1974) Spurious regressions in econometrics, *Journal of Econometrics*, 2.2, 111–120.

Gremeaux, V., Gayda, M., Lepers, R, Sosner, P., Juneau, M. and Nigam, A. (2012) Exercise and longevity, *Maturitas*, 73, 312-17.

Headey, B.W. (2006) Subjective well-being: revisions to dynamic equilibrium theory using national panel data and panel regression methods, *Social Indicators Research*, 79, 369-403.

Headey, B.W. (2008) Life goals matter to happiness: A revision of set-point

theory, *Social Indicators Research*, 86, 213-31.

Headey, B.W., Hoehne, G. and Wagner G.G. (2014) Does religion make you healthier and longer lived? Evidence for Germany, *Social Indicators Research*, 119, 1335-61.

Headey, B.W., Muffels, R.J.A. and Wagner, G.G. (2010) Long-running German panel survey shows that personal and economic choices, not just genes, matter for happiness, *Proceedings of the National Academy of Sciences*, 107.42, 17922-17926 (Oct. 19).

Headey, B.W., Muffels, R.J.A. and Wagner, G.G. (2013) Choices which change life satisfaction: Similar results for Australia, Britain and Germany, *Social Indicators Research*, 112, 725-48.

Headey, B.W. and Muffels, RJA (2015) Towards a theory of medium term Life Satisfaction: Two-way causation partly explains persistent satisfaction or dissatisfaction, *Social Indicators Research*, 124.2.

Published online 14 October 2015. DOI 10.1007/s11205-015-1146-8.

Headey, B.W., Schupp, J., Tucci, I. and Wagner, G.G. (2010) Authentic happiness theory supported by impact of religion on life satisfaction: A longitudinal analysis with data for Germany, *Journal of Positive Psychology*, 5, 73-82.

Headey, B.W., Veenhoven, R. and Wearing, A.J. (1991) Top-down versus bottom-up theories of subjective well-being, *Social Indicators Research*, 24, 81-100.

Henderson, S., Byrne, D.G. and Duncan-Jones, P. (1981) *Neurosis and the Social Environment*. New York, Academic Press.

Kessler, R.C. and Greenberg, D.F. (1981) *Linear Panel Analysis*. New York, Academic Press.

Kuskova, V.A. (2011) A longitudinal analysis of the relationship between life satisfaction and employee volunteerism, *Academy Of Management Proceedings*, 1, 1-6.

Lance, C. E., Mallard, A.G. and Michalos, A.C. (1995) Tests of the causal directions of global-life facet satisfaction relationships, *Social Indicators Research*, 34, 69–92.

Lucas, R.E. (2008) Personality and subjective well-being in M. Eid and R.J. Larsen eds. *The Science of Subjective Well-Being*. New York, Guilford Press, pp. 171-94.

Lucas, R.E., Clark, A.E., Georgellis, Y. and Diener, E. (2003)

Reexamining

adaptation and the set point model of happiness: Reactions to change in marital

status, *Journal of Personality and Social Psychology*, 84, 527-39.

Luhmann, M. and Eid, M. (2009) Does it really feel the same? Changes in life satisfaction following repeated life events, *Journal of Personality and Social Psychology*, 97, 363-81.

Luhmann, M., Hoffman, W., Eid, M. and Lucas, R.E. (2012) Subjective well-being and adaptation to life events: A meta-analysis, *Journal of Personality and Social Psychology*, 102, 592-615.

Lykken, D. and Tellegen, A. (1996) Happiness is a stochastic phenomenon,

Psychological Science, 7, 186-89.

Lynn, P. (2006) Quality Profile: BHPS Version 2.0: Waves 1 to 13 1991-2003.

University of Essex, Institute for Social Research.

Lyubomirsky, S., Sheldon, K.M. and Schkade, D. (2005) Pursuing happiness:

The architecture of sustainable change, *Review of General Psychology*, 9, 111-

31.

Mathison, L., Andersen, M.H., Veenstra, M., Wahl, A.K., Hanestad, B.R. and Fosse, E. (2007) Quality of life can both influence and be an outcome of general health perceptions after heart surgery, *Health and Quality of Life Outcomes*, 5:27 doi:1186/1477-7525-5-27.

Mehnert, T., Kraus, H.H., Nadler, R. and Boyd, M. (1990) Correlates of life satisfaction in those with a disabling condition, *Rehabilitation Psychology*, 35, 3-17.

Meier, S. and Stutzer, A. (2004) Is volunteering rewarding in itself? IZA Discussion Paper No 1045 (March) IZA, Bonn.

Nagazato, N., Schimmack, U., and Oishi, S. (2011) Effect of changes in living conditions on well-being: a prospective top-down bottom-up model, *Social Indicators Research*, 100, 115-35.

Nickerson, C., Schwarz, N., Diener, E. and Kahneman, D. (2003) Zeroing in on the dark side of the American dream: A closer look at the negative consequences of the goal for financial success, *Psychological Science*, 14, 531-36.

Pearl, J. *Causality: Models, Reasoning and Inference*. Cambridge, Cambridge University Press, 2009.

Scherpenzeel, A. and Saris, W.E. (1996) Causal direction in a model of life satisfaction: the top-down/bottom-up controversy, *Social Indicators Research*, 38, 161-180.

Schwarze, J., Andersen, H and Silke, A. (2000) Self-rated health and changes in self-rated health as predictors of mortality – first evidence from the German panel data, DIW Discussion Paper No. 203, Berlin, DIW.

Sheldon, K.M. and Lucas, R.E. (2014) *The Stability of Happiness*.

Amsterdam, Elsevier.

StataCorp (2013) *Structural Equation Modelling Reference Manual*,

Release 13. College Station Texas, Stata Press.

Thoits, P.A. and Hewitt, L.N. (2001) Volunteer work and well-being,

Journal of Health and Social Behavior, 42, 115-31.

Ware, J.E., Snow, K and Kosinski, M. (2000) *SF-36 Health Survey:*

Manual and Interpretation Guide. Lincoln, Rhode Island, QualityMetric

Inc.

Watson, N. and Wooden, M. (2004) Assessing the quality of the HILDA

Survey

Wave 2 data. Melbourne, Melbourne Institute of Applied Economic and

Social

Research.

Wilkins, A.S. (2014) To lag or not to lag: Re-evaluating the use of lagged

dependent variables in regression analysis. Working Paper, Stanford

University Department of Political Science. Downloaded July 4, 2014.

Wooldridge, J.M. (2010) *Econometric Analysis of Cross-Section and Panel Data*. Cambridge, Mass., MIT Press, 2nd edition.

TABLES

(the 2 Figures are in separate files)

Table 1

Within-Person (Fixed Effects) Regressions Relating LS_t to LS_{t+1} ,
 LS_{t-1} etc: Metric Coefficients with Standard Errors in Parentheses.
 Results for Australia (N=55632), Britain (N=47649) & Germany
 (N=59024)^a

Dependent variable	Australia: independent variable LS_t	Britain: independent variable LS_t	Germany: independent variable LS_t
LS_{t+1}	0.100*** (0.004)	0.253*** (0.004)	0.213*** (0.004)
LS_{t-1}	0.105*** (0.004)	0.232*** (0.004)	0.216*** (0.004)
LS_{t+2}	0.015*** (0.004)	0.036*** (0.005)	0.114 *** (0.004)
LS_{t-2}	0.017*** (0.004)	0.031*** (0.004)	0.116*** (0.004)
LS_{t+3}	-0.045*** (0.004)	-0.022*** (0.005)	0.053 *** (0.004)
LS_{t-3}	-0.050*** (0.004)	-0.019*** (0.004)	0.054 *** (0.004)

a. N = person-years.

*** significant at the 0.001 level

Table 2: Australia

Two-Way Causal Links Between LS and Health SF-36 (N=41966)^c

	<i>Metric ML estimates^a</i>	<i>Standardized ML estimates^a</i>	<i>Measures of model fit^b</i>	
Health->LS (BU link)	0.005*** (0.000)	0.073*** (0.004)	LR Chi-square (df=113)	5508.803***
LS->Health (TD link)	0.279*** (0.040)	0.018*** (0.003)	RMSEA	0.034
Correlated error (SP link)	2.030*** (0.083)	0.172*** (0.006)	SRMR	0.031
			CFI	0.982
			TLI	0.974

a. Robust standard errors in parentheses. *** significant at the 0.001 level.

b. Model stability= 0.0004; all eigenvalues inside unit circle (Bentler and Freeman, 1983).

c. N = person-years.

Table 2: Britain

Two-Way Causal Links Between LS and Self-Rated Health (N=36647)^c

	<i>Metric ML estimates^a</i>	<i>Standardized ML estimates^a</i>	<i>Measures of model fit^b</i>	
Health->LS (BU link)	0.083*** (0.002)	0.062*** (0.002)	LR Chi-square (df=128)	10447.82***
LS->Health (TD link)	0.052*** (0.001)	0.069 (0.002)	RMSEA	0.035
Correlated error (SP link)	0.091*** (0.002)	0.156*** (0.004)	SRMR	0.033
			CFI	0.971
			TLI	0.959

a. Robust standard errors in parentheses. *** significant at the 0.001 level.

b. Model stability= 0.0003; all eigenvalues inside unit circle (Bentler and Freeman, 1983).

c. N = person-years.

Table 2: Germany

Two-Way Causal Links Between Self-Rated Health and LS (N= 56729)^a

	<i>Metric ML estimates</i>	<i>Standardized ML estimates</i>	<i>Measures of model fit^b</i>	
Health _t -->LS _{t+1} (BU link)	0.155 *** (0.004)	0.079 *** (0.002)	LR Chi-square (df=128)	8406.45***
LS _t ->Health _{t+1} (TD link)	0.029 *** (0.001)	0.057 *** (0.002)	RMSEA	0.034
Correlated error (SP link)	0.182 *** (0.003)	0.233 *** (0.004)	SRMR	0.033
			CFI	0.975
			TLI	0.964

a. N=person years

b. Model stability=0.000; all eigenvalues inside unit circle (Bentler and Freeman, 1983).

*** significant at the 0.001 level

Table 3: Australia

Two-Way Causal Links Between LS and Social Support (N=43261)^c

	<i>Metric ML estimates^a</i>	<i>Standardized ML estimates^a</i>	<i>Measures of model fit^b</i>	
Social support ->LS (BU link)	0.060*** (0.003)	0.085*** (0.004)	LR Chi-square (df=122)	7402.818***
LS->Social support (TD link)	0.079*** (0.005)	0.056*** (0.003)	RMSEA	0.037
Correlated error (SP link)	0.222*** (0.009)	0.167*** (0.006)	SRMR	0.035
			CFI	0.972
			TLI	0.959

a. Robust standard errors in parentheses. *** significant at the 0.001 level.

b. Model stability=0.0003; all eigenvalues inside unit circle (Bentler and Freeman, 1983).

c. $N = \text{person-years}$.

Table 3: Britain

Two-Way Causal Links Between LS and Social Participation (N=36572)^c

	<i>Metric ML estimates^a</i>	<i>Standardized ML estimates^a</i>	<i>Measures of model fit^b</i>	
Social participation ->LS (BU link)	0.003*** (0.003)	0.018*** (0.002)	LR Chi-square (df=136)	8658.74***
LS->Social participation (TD link)	0.012*** (0.001)	0.023*** (0.002)	RMSEA	0.030
Correlated error (SP link)	0.019*** (0.002)	0.042*** (0.004)	SRMR	0.026
			CFI	0.974
			TLI	0.962

a. Robust standard errors in parentheses. *** significant at the 0.001 level.

b. Model stability=0.0002; all eigenvalues inside unit circle (Bentler and Freeman, 1983).

c. N = person-years.

Table 3: Germany

Two-Way Causal Links Between Social Participation and LS (N=56593)^a

	<i>Metric ML estimates</i>	<i>Standardized ML estimates</i>	<i>Measures of model fit^b</i>	
Social participation _t -->LS _{t+1} (BU link)	0.126 *** (0.007)	0.032 *** (0.002)	LR Chi-square (df=136)	5455.33***
LS _t ->Social participation _{t+1} (TD link)	0.008 *** (0.000)	0.030 *** (0.002)	RMSEA	0.026
Correlated error (SP link)	0.010 *** (0.002)	0.026 *** (0.004)	SRMR	0.021
			CFI	0.983
			TLI	0.975

a. N=person years

b. Model stability=0.000; all eigenvalues inside unit circle (Bentler and Freeman, 1983).

*** significant at the 0.001 level

Table 4: Australia

Two-Way Causal Links Between LS and Exercise (N=43927)^c

	<i>Metric ML estimates^a</i>	<i>Standardized ML estimates^a</i>	<i>Measures of model fit^b</i>	
Exercise->LS (BU link)	0.028*** (0.003)	0.031*** (0.003)	LR Chi-square (df=120)	5167.150***
LS->Exercise (TD link)	0.017*** (0.004)	0.016*** (0.003)	RMSEA	0.031
Correlated error (SP link)	0.060*** (0.007)	0.050*** (0.006)	SRMR	0.030
			CFI	0.974
			TLI	0.961

a. Robust standard errors in parentheses. *** significant at the 0.001 level.

b. Model stability=0.0002; all eigenvalues inside unit circle (Bentler and Freeman, 1983).

c. N = person-years.

Table 4: Britain

Two-Way Causal Links Between LS and Exercise (N=36732)^c

	<i>Metric ML estimates^a</i>	<i>Standardized ML estimates^a</i>	<i>Measures of model fit^b</i>	
Exercise->LS (BU link)	0.029*** (0.001)	0.035*** (0.002)	LR Chi-square (df=136)	9478.91***
LS->Exercise (TD link)	0.009*** (0.001)	0.007*** (0.001)	RMSEA	0.032
Correlated error (SP link)	0.010*** (0.002)	0.024*** (0.004)	SRMR	0.031
			CFI	0.987
			TLI	0.981

a. Robust standard errors in parentheses. *** significant at the 0.001 level.

b. Model stability=0.0007; all eigenvalues inside unit circle (Bentler and Freeman, 1983).

c. N = person-years.

Table 4: Germany

Two-Way Causal Links Between Exercise and LS (N=55653)^a

	<i>Metric ML estimates</i>	<i>Standardized ML estimates</i>	<i>Measures of model fit^b</i>	
Exercise _t -->LS _{t+1} (BU link)	0.047 *** (0.003)	0.034 *** (0.002)	LR Chi-square (df=136)	5102.34***
LS _t ->Exercise _{t+1} (TD link)	0.010 *** (0.001)	0.014 *** (0.001)	RMSEA	0.026
Correlated error (SP link)	0.024 *** (0.004)	0.027 *** (0.004)	SRMR	0.022
			CFI	0.988
			TLI	0.982

a. N=person-years.

b. Model stability=0.000; all eigenvalues inside unit circle (Bentler and Freeman, 1983).

*** significant at the 0.001 level

Table 5: Australia

Two-Way Causal Links: LS with Health SF-36, Social Support and
Exercise (N=40211)^c

	<i>Metric ML estimates^a</i>	<i>Standardized ML estimates^a</i>	<i>Measures of model fit^b</i>	
<i>Health</i>				
Health->LS (BU link)	0.003*** (0.000)	0.043*** (0.002)	LR Chi-square (df=302)	20117.096***
LS->Health (TD link)	0.347*** (0.041)	0.023*** (0.002)	RMSEA	0.039
Correlated error (SP link)	1.974*** (0.084)	0.170*** (0.005)	SRMR	0.052
<i>Social support</i>			CFI	0.963
Social support ->LS (BU link)	0.048*** (0.003)	0.069*** (0.002)	TLI	0.948

LS->Social support (TD link)	0.082*** (0.005)	0.058*** (0.002)		
Correlated error (SP link)	0.222*** (0.009)	0.170*** (0.005)		
<i>Exercise</i>				
Exercise- >LS (BU link)	0.008** (0.002)	0.009** (0.002)		
LS- >Exercise (TD link)	0.019*** (0.004)	0.017*** (0.002)		
Correlated error (SP link)	0.055*** (0.007)	0.046*** (0.005)		

a. Robust standard errors in parentheses. ***significant at the 0.001 level.

**significant at the 0.01 level.

b. Model stability=0.0004; all eigenvalues inside unit circle (Bentler and Freeman, 1983).

c. N = person-years.

Table 5: Britain

Two-Way Causal Links: LS with Self-Rated Health, Social Participation
and Exercise (N=36630)^c

	<i>Metric ML estimates^a</i>	<i>Standardized ML estimates^a</i>	<i>Measures of model fit^b</i>	
<i>Health</i>				
Health->LS (BU link)	0.081*** (0.002)	0.061*** (0.002)	LR Chi- square (df=300)	16791.64***
LS->Health (TD link)	0.052*** (0.001)	0.070*** (0.002)	RMSEA	0.029
Correlated error (SP link)	0.091*** (0.002)	0.155*** (0.004)	SRMR	0.038
<i>Social participation</i>			CFI	0.983
Social participation ->LS (BU link)	0.041*** (0.003)	0.022*** (0.002)	TLI	0.976

LS->Social participation (TD link)	0.013*** (0.001)	0.023*** (0.002)		
Correlated error (SP link)	0.019*** (0.002)	0.044*** (0.004)		
<i>Exercise</i>				
Exercise->LS (BU link)	0.011*** (0.001)	0.013*** (0.002)		
LS->Exercise (TD link)	0.009*** (0.001)	0.007*** (0.001)		
Correlated error (SP link)	0.008*** (0.002)	0.019*** (0.004)		

a. Robust standard errors in parentheses ***significant at the 0.001 level.

b. Model stability=0.0004; all eigenvalues inside unit circle (Bentler and Freeman, 1983).

c. N = person-years.

Table 5: Germany

Two-Way Causal Links: LS Related to Self-Rated Health, Social Support
and Exercise (N=55512)^a

	<i>Metric ML estimates</i>	<i>Standardized ML estimates</i>	<i>Measures of model fit^b</i>	
<i>Health</i>			LR Chi-square (df=332)	13732.53***
Health _t ->LS _{t+1}	0.148 *** (0.004)	0.078 *** (0.002)	RMSEA	0.027
LS _t ->Health _{t+1}	0.029 *** (0.001)	0.057 *** (0.002)	SRMR	0.033
Correlated error (SP link)	0.182 *** (0.003)	0.234 *** (0.004)	CFI	0.982
<i>Social participation</i>			TLI	0.975
Social participation _t -->LS _{t+1}	0.097*** (0.010)	0.022 *** (0.002)		
LS _t ->Social	0.005***	0.014 ***		

participation _{t+1}	(0.000)	(0.001)		
Correlated error	0.007*** (0.002)	0.024 *** (0.004)		
<i>Exercise_t</i>				
Exercise _t --> LS _{t+1}	0.016 *** (0.003)	0.011 *** (0.002)		
LS _t ->Exercise _{t+1}	0.010 *** (0.001)	0.014 *** (0.001)		
Correlated error (SP link)	0.022 *** (0.004)	0.025 *** (0.004)		

a. N=person years. b. Model stability=0.000; all eigenvalues inside unit circle (Bentler and Freeman, 1983). *** significant at the 0.001 level

Table 6: Australia

Two-Way Causal Links: LS with Job Satisfaction and Partner

Satisfaction (N=15848)^c

	<i>Metric ML estimates^a</i>	<i>Standardized ML estimates^a</i>	<i>Measures of model fit^b</i>	
<i>Job satisfaction</i>				
Job satis.->LS (BU link)	0.041*** (0.005)	0.055*** (0.004)	LR Chi-square (df=196)	1923.755***
LS->Job satis. (TD link)	0.163*** (0.011)	0.080*** (0.004)	RMSEA	0.024
Correlated error (SP link)	0.308*** (0.013)	0.288*** (0.007)	SRMR	0.031
<i>Partner satisfaction</i>			CFI	0.985
Partner satis.->LS	0.051*** (0.004)	0.080*** (0.004)	TLI	0.978

(BU link)				
LS->Partner satis. support (TD link)	0.109*** (0.011)	0.067*** (0.004)		
Correlated error (SP link)	0.228*** (0.014)	0.204*** (0.011)		

a. Robust standard errors in parentheses. *** significant at the 0.001 level.

b. Model stability=0.0002; all eigenvalues inside unit circle (Bentler and Freeman, 1983).

c. N = person-years.

Table 6: Britain

Two-Way Causal Links: LS with Job Satisfaction and Partner

Satisfaction (N=19428)^c

	<i>Metric ML estimates^a</i>	<i>Standardized ML estimates^a</i>	<i>Measures of model fit^b</i>	
<i>Job satisfaction</i>				
Job satis.- >LS (BU link)	0.045*** (0.002)	0.057*** (0.003)	LR Chi-square (df=220)	5686.485***
LS->Job satis. (TD link)	0.118*** (0.004)	0.093*** (0.003)	RMSEA	0.027
Correlated error (SP link)	0.279*** (0.005)	0.327*** (0.005)	SRMR	0.027
<i>Partner satisfaction</i>			CFI	0.983
Partner satis.	0.012***	0.025***	TLI	0.976

->LS (BU link)	(0.002)	(0.002)		
LS->Partner satis. (TD link)	0.081*** (0.004)	0.039*** (0.002)		
Correlated error (SP link)	0.254*** (0.006)	0.244*** (0.005)		

a. Robust standard errors in parentheses. *** significant at the 0.001

level.

b. Model stability=0.0003; all eigenvalues inside unit circle (Bentler and Freeman, 1983).

c. N = person-years.

Table 6: Germany

Two-Way Causal Links: LS with Job Satisfaction and Satisfaction with Family Life (N=5622)^c

	<i>Metric ML estimates^a</i>	<i>Standardized ML estimates^a</i>	<i>Measures of model fit^b</i>	
<i>Job satisfaction</i>				
Job satis.- >LS (BU link)	0.048*** (0.005)	0.060*** (0.006)	LR Chi-square (df=220)	1102.563***
LS->Job satis. (TD link)	0.224*** (0.010)	0.177*** (0.008)	RMSEA	0.027
Correlated error (SP link)	0.537*** (0.027)	0.278*** (0.012)	SRMR	0.029
<i>Satisfaction with family life</i>			CFI	0.978

Family satis. ->LS (BU link)	0.054*** (0.005)	0.066*** (0.007)	TLI	0.969
LS->Family satis. (TD link)	0.124*** (0.008)	0.103*** (0.007)		
Correlated error (SP link)	0.495*** (0.024)	0.281*** (0.012)		

a. Robust standard errors in parentheses. *** significant at the 0.001

level.

b. Model stability=0.0003 ; all eigenvalues inside unit circle (Bentler and Freeman, 1983).

c. N = person-years.