Soc Indic Res
DOI 10.1007/s11205-009-9516-8

Environmental Influences on Well-Being: A Dyadic Latent Panel Analysis of Spousal Similarity

Ulrich Schimmack · Richard E. Lucas

Accepted: 23 August 2009

Abstract This article uses dyadic latent panel analysis (DLPA) to examine environmental influences on well-being. DLPA requires longitudinal dyadic data. It decomposes the observed variance of both members of a dyad into a trait, state, and an error component. Furthermore, state variance is decomposed into initial and new state variance. Total observed similarity between members of a dyad is decomposed into trait similarity, initial state similarity, new state similarity, and error similarity. Dyadic similarity in new state variance reveals that both members of a dyad change in the same direction, which is a strong indication of environmental effects. DLPA is used to examine environmental influences on life satisfaction and domain satisfaction based on 22 annual assessments of married couples in the German Socio-Economic Panel Study (N's = 607–740). The results show high similarity in new state variance for life satisfaction and objectively identical domains (household income, housing), and less similarity for objectively less similar domains (recreation, health). This finding provides strong evidence for environmental influences on well-being. In addition, the results show high trait similarity. The implications of the latter finding for interpretation of behavioral genetics studies of well-being are discussed.

Keywords Well being · Satisfaction · Environment · Longitudinal · Dyadic · Couples · Spouses · Heritability

1 Introduction

One of the most interesting and hardest questions in the social sciences is the origin of individual differences. Why are some individuals more or less conscientious, intelligent, healthy, or happy than others? Investigating this important question is difficult because

U. Schimmack (✉)
Department of Psychology, University of Toronto, Mississauga (UTM), 3359 Mississauga Road North, Mississauga, ON L5L 1C6, Canada
e-mail: uli.schimmack@utoronto.ca

R. E. Lucas
Michigan State University, East Lansing, MI, USA

Published online: 27 September 2009

studies of naturally occurring individual differences rely on correlations to make inferences about causality. Social scientists are well aware of the logical fallacy cum hoc ergo propter hoc to infer causality from a single correlation. It is important to note, however, that correlations provide useful information for causal theories, and that a rigorous program of correlational research can reveal causal processes. This article introduces Dyadic Latent Panel Analysis to examine whether environmental factors *cause* individual differences in well-being. After an introduction of the causal model and its implementation in structural equations, Dyadic Latent Panel Analysis is applied to dyadic longitudinal panel data to examine the influence of environmental factors on well-being.

1.1 Inferring Causality from Correlations: Monozygotic Twins Reared Apart

Behavioral genetics research on twins provides the best demonstration of the power of correlational research to demonstrate causality, even when the actual causal variable (i.e., genes) was never observed. Only 40 years ago, the prevalent view in the social sciences was that individual differences are predominantly determined by situational factors that may change rapidly from moment to moment. Today, most social scientists acknowledge that individual differences have considerable stability over long periods of time and that genetic differences explain a substantial portion of individual differences in personality, intelligence, and attitudes (Mischel 1968; Ross and Nisbett 1991; Tesser 1993).

Twin studies allow inferences about causality from correlations because correlations can provide important information about causality. After all, some causal process must produce an observed correlation between two variables (Neale and Cardon 1992). At the most abstract level, four causal processes that are not mutually exclusive can produce a correlation between variables A and B: A causes B (1), B causes A (2), or a third variable C causes A (3) and B (4). A single observed correlation is insufficient to estimate the four parameters of the causal model. However, additional assumptions or empirical observations can reduce the number of parameters that need to be estimated. For example, a single correlation between MZ twins reared apart is sufficient to determine causality because twins reared apart cannot influence each other. As a result, it is possible to rule out a causal influence of Twin A on Twin B and vice versa. Furthermore, there is no valid distinction between Twin A and Twin B. As a result, the causal effect of the third variable, C, is equivalent for both. Given $A \rightarrow B = 0$, and $B \rightarrow A = 0$, and $C \rightarrow A = C \rightarrow B$, the causal model reduces from four parameters to one parameter, and it becomes possible to estimate the unknown causal effect based on a single observed correlation. Furthermore, the single parameter can be interpreted as the effect of genes because the only common causal factors between twins reared apart are their genes.

1.2 Limitations of Twin Studies with Twins Reared Together

The example of monozygotic twins reared apart is exceptional. It is more difficult to make inferences about causality when twins share both genetic dispositions and environmental influences (Neale and Cardon 1992). As noted by Eaves et al. (1989) “there are effects that cannot be resolved with data on twins by themselves, including the effects of assortment and the shared environment.” (p. 113).

Another limitation of twin studies is that they are more suitable to demonstrate genetic than environmental influences on individual differences. The reason is that estimates of environmental influences combine three distinct factors. One factor, the shared environment, produces similarity between twins. For example, parental divorce may have negative effects on the well-being of both twins. The other two factors make twins dissimilar. These two factors are non-shared environment and measurement error. For many individual differences the shared environment between twins is the smallest factor and often makes a negligible contribution. Indeed, twin studies of well-being in adulthood consistently show no effect of the shared environment (Roysamb et al. 2002; Stubbe et al. 2005; Weiss et al. 2008). Thus, estimates of environmental influences are typically limited to influences of the non-shared environment and measurement error.

A major limitation of the standard twin study is that it cannot distinguish between non-shared environmental factors and measurement error. Thus, measurement error may produce spurious evidence for environmental influences (Schmidt and Hunter 1996). For example, Stubbe et al. (2005) estimated that 38% of the total variance in life satisfaction was heritable, and that the remaining 62% of the variance was caused by unshared environmental factors and measurement error. Without a precise estimate of the amount of measurement error, the estimate for environmental factors is inflated, and the estimate for heritability is attenuated to an unknown degree. Indeed, longitudinal twin studies of well-being found that most of the stable variance in subjective well-being was heritable, whereas environmental influences and measurement error account mainly for the unstable variance components (Lykken and Tellegen 1996; Nes et al. 2006). This finding may suggest that environmental influences are transitory, but it may also indicate that a considerable amount of the unstable variance is measurement error.

Another limitation of previous twin studies is uncertainty about the genetic similarity of dizygotic twins and siblings when the genetic similarity between parents is unknown. All studies work with the simple assumption that dizygotic twins share 50% of their genes. However, this assumption assumes that spouses are not genetically similar, but some previous studies have suggested that spouses are genetically similar (Tambs and Moum 1992). As a result, estimates of the additive genetic influences on well-being may be inflated.

In short, twin studies make an important contribution to the study of genetic versus environmental influences on well-being. However, like all other correlational studies, the standard twin design relies on assumptions to draw causal inferences from correlational data, and rigorous tests of these assumptions are lacking.

1.3 Inferring Causality from Correlations: Spousal Similarity

The study of dyads that are not genetically related overcomes some of the problems inherent in the study of dyads that share genetic and environmental influences in traditional behavioral genetics studies. This article focuses on spousal similarity, but the discussion of spousal similarity can be generalized to other genetically unrelated dyads (e.g., roommates, co-workers, neighbors, etc.). Studies of spouses can separate environmental influences from measurement error because environmental influences on spouses are revealed as shared environment effects, which are not confounded with measurement error. For example, if household income were a positive predictor of well-being, then spouses should have similar levels of well-being because spouses share the same household income.

It is possible to consider spouses as a special case of a traditional behavioral genetics study. In a typical behavioral genetics study of biologically related dyads, genetic similarity is set to a fixed value based on the degree of genetic relatedness (e.g., 1 for MZ twins, .5 for DZ twins, .25 for grandparent-grandchild). For spouses the genetic similarity due to genetic relatedness is zero, under the standard assumption of random mating. The ability to fix genetic similarity due to genetic relatedness to zero provides the ability to reveal environmental influences on individual differences with fewer assumptions than studies that have to separate dyadic similarity into a genetic and a shared environmental factor.

In this regard, the advantage of studying spouses for uncovering environmental factors is similar to the advantage of studying mono-zygotic twins reared apart for uncovering genetic factors. For mono-zygotic twins reared apart the environmental influences can be set to zero. For spouses the genetic similarity can be set to zero. In both cases, the ability to fix a parameter to a known value reduces uncertainty about the value of other parameters in the model. The ability to fix genetic relatedness to zero makes studies of spousal similarity akin to adoption studies, in which genetic relatedness is also zero. However, adoption studies have a number of other problems. First, the well-being of parents and their children is likely to be influenced by different factors due to the age difference. In contrast, most spouses are similar in age. Second, adoptions are unusual which raises questions about the generalizability of the findings. In contrast, marriage is still a common practice even in modern Western societies. Thus, studies of spouses can be generalized to a large percentage of the adult population.

Given the lack of evidence for shared environment effects on well-being of adult twins, it is important to emphasize that shared environment effects for spouses are fundamentally different from shared environment effects for adult twins (Caspi et al. 1992). In a study of adult twins, shared environment captures mostly past environmental factors (the shared rearing environment; Caspi et al. 1992). In contrast, spousal similarity reveals current environmental influences (e.g., household income, social influences of spouses on each other, etc.; i.e., the shared marital environment; Caspi et al. 1992). Thus, shared environment factors for spouses are to a large extent part of the unshared environment variance in twin studies.

One reason for the failure to find shared environment effects in twin studies may be that well-being is influenced primarily by recent environmental factors. Consistent with this hypothesis, twin studies often find a decrease in shared environment effects during adolescence (McCartney et al. 1990). This finding suggests that recent environmental factors that are not shared with a twin become more important over time, whereas the effects of shared past experiences (e.g., parenting styles) become weaker.

The difference between shared environment with a twin and shared environment with a spouse is illustrated in a study of marital satisfaction in female twins and their husbands (Spotts et al. 2004). The study found that the shared environment with a co-twin had no influence on marital satisfaction (i.e., genes were sufficient to explain twin similarity in marital satisfaction). At the same time, shared environmental factors between spouses, reflected in high spousal similarity, made a substantial contribution to marital satisfaction. In other words, the current shared environment with a spouse influenced marital satisfaction, but the past shared-environment with a twin did not. It is possible that these findings generalize to other indicators of well-being because marital satisfaction contributes to life-satisfaction (Headey et al. 1991).

1.4 Assortative Mating

Spousal similarity is not sufficient to prove environmental influences on individual differences. The reason is that spousal similarity may be due to assortative mating; that is, the tendency to select a spouse with similar attributes. For example, spouses are similar in height and individual differences in height are highly heritable (Silventoinen et al. 2003). Thus, spousal similarity in height erroneously suggests an environmental influence on height during adulthood, if one neglects assortative mating as an alternative causal mechanism. However, it is equally problematic to assume that spousal similarity is caused by assortative mating, when it is in fact due to shared environmental influences. "Incorrect or untested assumptions about assortative mating for example, will usually lead to incorrect conclusions about genetic architecture and environmental transmission" (Heath and Eaves 1985, p. 16).

Despite the critical importance of separating environmental influences and assortative mating as causes of spousal similarity, this issue has been neglected in empirical studies of well-being. The main reason is probably that most dyadic studies are cross-sectional, and cross-sectional studies make it virtually impossible to distinguish environmental influences and assortative mating. Longitudinal dyadic studies overcome this limitation (Schimmack and Lucas 2007). The reason is that assortative mating and shared environment effects make different predictions about the time course of spousal similarity. Assortative mating predicts spousal similarity at the beginning of the marital relationship. It predicts spousal similarity during later stages of the relationship only to the extent that the factors that produced initial similarity remain stable over time. In contrast, changes in the characteristics that produced assortment would reduce spousal similarity (Caspi et al. 1992). For example, spouses tend to assort on physical attractiveness (Feingold 1988). However, physical attractiveness changes gradually over time (Alley 1993; Zebrowitz et al. 1993). Thus, an assortative mating model predicts that spousal similarity in physical attractiveness decreases over time. To illustrate, assume that spousal similarity in attractiveness at the beginning of marriage is $r = .4$, and that physical attractiveness has a 10-year stability of .8. Due to the changes in attractiveness over time, spousal similarity in attractiveness after 10 years of marriage would be only $r = .25$ (i.e., $.4 \times .8 \times .8$). Effects of the shared environment make a different prediction. In the previous example, spousal similarity in the factors that produce changes in physical attractiveness would add to the .25 similarity that remains due to initial assortment. Whether spousal similarity in attractiveness after 10-years is higher or lower than initial similarity ($r = .4$) depends on the effect size of spousal similarity in the new variance that produces changes in attractiveness over time.

In general, a longitudinal study of spousal similarity can be used to distinguish four factors of spousal similarity: (a) spousal similarity in traits that do not change over the observed time interval, (b) initial spousal similarity in states that change over time, (c) spousal similarity in new state variance that produces changes in state variance over time, and (d) shared measurement error between spouses. Of these four components, spousal similarity in new state variance that is not present at the beginning of marriage provides strong evidence for environmental influences because it cannot be attribute to assortative mating (Schimmack and Lucas 2007).

It is possible that spousal similarity in new state variance reflects pure environmental effects or gene $\times$ environment interactions. For example, all spouses may respond with an increase in well-being in response to an increase in household income. Alternatively, the effect of an increase in household income on spouse's well-being may be moderated by genetic dispositions that influence materialistic values. In this article, we do refer to all effects that involve environment as environmental effects and do not distinguish between main effects of environment or interaction effects of environment with genetic dispositions. Using this broad definition of environmental effects, spousal similarity in new state variance that emerges over the course of marriage is necessary and sufficient to demonstrate environmental effects on well-being. In other words, longitudinal dyadic data provide the opportunity to make strong inferences about causality on the basis of correlations and without prior knowledge of the environmental factors that influence well-being.

Dyadic longitudinal studies of spouses have an advantage over longitudinal twin designs to demonstrate environmental effects. The reason is that shared environment effects in twin studies tend to be negligible. As a result, environmental influences are limited to the unique variance of each twin and changes in the unique variance can reflect true changes in well-being or mere random measurement error. In contrast, demonstrating similar changes in dyads that share environmental factors eliminate random measurement error as a potential causal factor.

1.5 Dyadic Latent Panel Analysis of Spousal Similarity

Several articles have made important contributions to the modeling of longitudinal data (Ehrhardt et al. 2000; Kenny and Zautra 1995; Ormel and Schaufeli 1991; Rudinger and Rietz 1998; Steyer et al. 1999). The main common element of these models is that they decompose the total variance of a measure into three components, a stable trait component that does not change over the observed time period, an unstable state component that changes over time, and a third component that reflects error variance. This model is typically applied to repeated measures of a single construct to examine the contribution of trait and state factors to stability and change of a single construct (Ehrhardt et al. 2000; Lucas and Donnellan 2007; Ormel and Rijsdijk 2000). An extension of the model to two constructs can be used to decompose the observed covariation between two constructs into covariations of the three variance-components (Kenny and Zautra 1995; Ormel and Schaufeli 1991). Dyadic Latent Panel Analysis uses this approach to examine the contribution of different variance components to observed spousal similarity.

Figure 1 provides a visual representation of dyadic latent panel analysis (DLPA). Like many structural equation models, DLPA starts with a measurement model that separates the observed variance (O) into two latent factors that represent reliable variance (R) and error variance (E). This decomposition is carried out for all repeated measures of wives (O_{w}) as well as all repeated measures of husbands (O_{H}). Although it is unusual to have a measurement model with a single observed variable, the longitudinal nature of the data makes it possible to separate reliable variance and error variance (Heise 1969; Kenny and Zautra 1995; Rudinger and Rietz 1998).

$$\text{Observed variance } (O) = \text{Reliable variance } (R) + \text{Error variance } (E) \tag{1}$$

The next assumption of DLPA is that reliable variance (R) is fully determined by two latent factors: a trait factor (T) that captures the influence of stable dispositions, and a state factor (S) that allows for changes over time. Again, this decomposition of reliable variance is applied to all repeated measures of wives (R_{W}) and husbands (R_{H}).

$$\text{Reliable variance } (R) = \text{State variance } (S) + \text{Trait variance } (T) \tag{2}$$

Dyadic latent panel analysis further distinguishes two components of state variance: old state variance (OS) that is shared with the state variance on the previous assessment and new state variance (N) that reflects valid changes in state variance from one assessment to the next. Figure 1 does not directly show OS. Rather, the amount of OS depends on the amount of state variance on the previous occasion (S_{n-1}) and the stability of state variance, which is represented by the autoregressive path coefficient of state variance at time $n-1$ to state variance at time n, β_{oldS}. OS equals the product of state variance on the previous occasion and the square of the autoregressive parameter, which reveals the amount of state variance that is explained by state variance on the previous occasion, $\mathrm{OS}_n = S_{n-1} \times \beta^2{}_{\mathrm{oldS}}$.

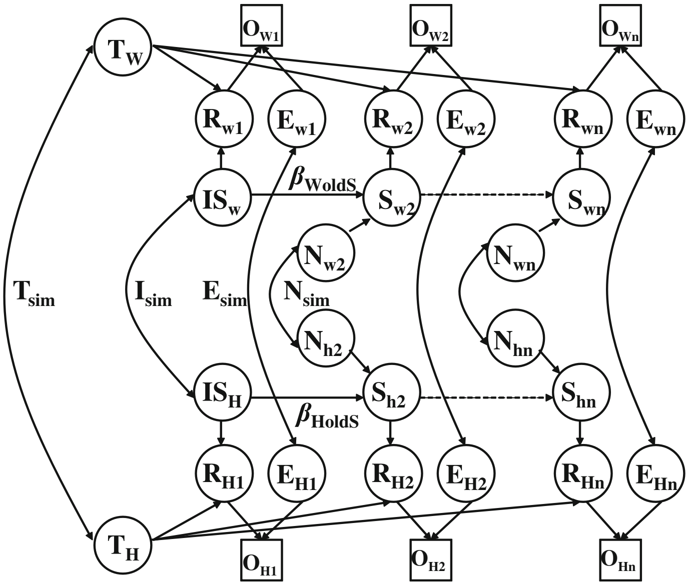

Fig. 1 Dyadic latent panel analysis of spousal similarity

$$\text{State variance } (S_n) = \text{Old state variance } (S_{n-1} \times \beta^2_{\text{oldS}}) + \text{New state variance } (N) \tag{3}$$

The decomposition of state variance into old and new variance is not possible for the first assessment. In Fig. 1, this variance is the initial state variance (IS).

The three equations imply that the model in Fig. 1 has five parameters to model stability and change of well-being for each spouse, a total of ten parameters.

$$O_{\mathrm{H}} = T_{\mathrm{H}} + \mathrm{IS}_{\mathrm{H}} \times \beta^2_{\mathrm{HoldS}} + N_{\mathrm{H}} + E_{\mathrm{H}} \tag{4a}$$

$$O_{\mathrm{W}} = T_{\mathrm{W}} + \mathrm{IS}_{\mathrm{W}} \times \beta^2_{\mathrm{WoldS}} + N_{\mathrm{W}} + E_{\mathrm{W}} \tag{4b}$$

In addition to these ten parameters, DLPA decomposes observed spousal similarity into similarity in four variance components:

1. Trait similarity (T_{sim}) is the correlation between the two trait factors T_{H} and T_{W}.
2. Initial state similarity (I_{sim}) is the correlation between the state factors on the first assessment IS_{H} and IS_{W}.
3. New state similarity (N_{sim}) is the similarity in the new state factors N_{H} and N_{W}.
4. Error similarity (E_{sim}) is the similarity in the error variances E_{H} and E_{W}.

For the current purpose of examining environmental influences on well-being, new state similarity is crucial because new state similarity reveals that spouses change in the same direction. The amount of new state variance over short time intervals is likely to be small because well-being is highly stable over short time-intervals (Ehrhardt et al. 2000; Lucas and Donnellan 2007). Thus, even high spousal similarity in new state variance can only make a small contribution to observed spousal similarity. However, these small effects on a single occasion accumulate over longer time intervals because state variance is fairly stable from year to year. Thus, it is also important to examine the cumulative amount of new state similarity. This can be done by examining spousal similarity in state variance on the last occasion. Over long time intervals most of the state variance at the end of the time interval is independent of initial state variance. Therefore, spousal similarity in state variance on the last occasion reveals the accumulated contribution of spousal similarity in new state variance.

The interpretation of the other three similarity components is more difficult. The nature of initial state similarity is ambiguous. It may reveal effects of assortment or prior environmental influence before the first assessment. Spousal similarity in trait variance may reveal genetic similarity due to assortment, but could also be due to stable shared environmental factors. Spousal similarity in the error variance also reveals changes, but these changes are limited to a single occasion and reflect temporary fluctuations or shared method variance.

In short, DLPA requires the estimation of 14 parameters: 5 parameters for wives' trait variance, initial state variance, new state variance, error variance, and stability of state variance, 5 parameters for husbands' trait variance, initial state variance, new state variance, error variance, and state stability, and 4 parameters for spousal similarity in trait variance, initial state variance, new state variance, and error variance. The actual number of parameters estimated in DLPA may be larger, depending on assumptions about stationarity of model parameters, as explained in the next section.

Dyadic latent panel analysis addresses one major concern in previous longitudinal dyadic studies with only two waves (Caspi et al. 1992; Tower and Kasl 1996). In these studies, it is difficult to distinguish spousal similarity in error variance from spousal similarity in new state variance. DLPA can separate these two sources and variance and estimate spousal similarity separately for each. Thus, DLPA provides more conclusive evidence for environmental influences.

1.6 Model Identification

Conditions for the identification of the DLST model are the same as for models with repeated assessments of a single individual (Cole et al. 2005; Kenny 1975). A key assumption of the model is that the four variances, trait, initial state, new state, and error, are independent. The assumption that error variance is independent of state and trait variance is common to structural equation models with a measurement model. The assumption of independence of state and trait variance is implied in the definition of trait and state variance as variance that is stable (trait) and variance that changes over time (state).

Another important assumption is stationarity (Kenny 1975). The stationarity assumption implies that the strength of causal processes that produce stability and change remains stationary over the different waves of assessment. This assumption implies that stability (and change) over the same time lag is the same. For example, the correlation between wave 1 and wave 2 is the same as the correlation between future waves with the same time interval (e.g., wave 5 and wave 6). Stationarity also implies that the amount of true variance remains constant from wave to wave. An additional assumption is that error variance remains constant across waves; that is, the reliability of assessments is stationary.

If all assumptions are made, DLPA estimates 14 parameters because parameters are constrained across waves. As a result, new waves do not add new information, but help to obtain more precise parameter estimates (Cole et al. 2005). However, additional waves can also be used to test stationarity assumptions (Kenny and Zautra 1995). If more than 8 waves are available, it is possible to assume stationarity for the first four waves, and for the second four waves, and to compare parameter estimates for the two sets of waves.

Significant differences between parameter estimates reveal violations of the stationarity assumption for the whole time interval. For example, stability may increase with age. In the present study with 22 waves of data, the waves were split into two halves of 11 waves to test stationarity.

1.7 Model Implementation

There are two reasons for the lack of applications of trait-state-error models to empirical data since Ormel and Schaufeli (1991) introduced the model. First, the model requires relatively large sample sizes and multiple waves to produce reliable estimates (Cole et al. 2005). Second, the stationarity assumption for the stat variance requires non-linear constraints, which were difficult to implement in linear structural equation programs (Rudinger and Rietz 1998; Schilling 2006). Recent advances in software development overcome this limitation. We used the software program MPLUS4.1 (Muthén and Muthén 2007) which makes it easy to implement non-linear constraints.[1] Users simply write the non-linear constraint in a single command line:

$$\text{Initial state variance} = \text{Initial state variance (IS)} \times \beta^2_{\text{oldS}} + \text{New variance}$$

1.8 Measurement of Well-Being

There exist numerous measures and definitions of well-being (Schimmack 2009). The present article focuses on cognitive measures of well-being or subjective evaluations of individuals' lives (Diener 1984). Cognitive measures of well-being ask respondents to evaluate their lives on a single evaluative continuum. The most commonly used measures are global life-satisfaction judgments. In addition, respondents may also evaluate specific aspects of their lives (e.g., work). Evaluations of specific domains are typically assessed with satisfaction ratings of life domains (Schimmack 2009). We focus on cognitive well-being for three reasons. First, some philosophers have proposed that life-satisfaction is the most adequate measure of well-being because it allows people to choose their own criteria to determine how well they are doing (Sumner 1996). Second, longitudinal panel studies have predominantly used cognitive measures of well being. As a result, our data set was limited to cognitive measures of well-being. Third, cross-sectional studies have demonstrated spousal similarity in life satisfaction (Bookwala and Schulz 1996; Schimmack and Lucas 2007; Tambs and Moum 1992). Thus, spousal similarity for cognitive measures of well-being is a robust phenomenon that requires further analysis. Indeed, past researchers disagree about the causes of spousal similarity in well-being. Tambs and Moum (1992) attributed spousal similarity in life satisfaction to assortative mating. The implicit implication of this conclusion is that spouses' shared environment has no influence on their well-being. In contrast, Bookwala and Schulz (1996) attributed spousal similarity in life-satisfaction to environmental influences, implicitly assuming that assortative mating does not contribute to spousal similarity in well-being. Dyadic longitudinal data are needed to test these alternative interpretations of cross-sectional data empirically.

Dyadic studies of domain satisfaction are rare in the well-being literature, with the exception of marital satisfaction, which shows high spousal similarity (Spotts et al. 2004). The investigation of spousal similarity in domain satisfaction is particularly interesting because life domains vary in the similarity of the objective factors that could influence

[1] An annotated Mplus syntax file is available from the first author upon request.

spouses' subjective evaluation of a life domain. Some life domains (e.g., household income, housing) are objectively identical for both spouses. Thus, high degrees of spousal similarity should be observed, if domain satisfaction were influenced by the common objective circumstances. In contrast, other domains are objectively less similar (e.g., health, recreation). For these domains, spousal similarity should be lower, if environmental factors influence satisfaction with these domains. For example, spouses should evaluate their health differently if one spouse is ill and the other spouse is healthy. Thus, an environmental model predicts higher spousal similarity for objectively more similar domains than objectively less similar domains. To test this hypothesis, DLPA was used to examine spousal similarity in satisfaction with two identical domains (household income, housing) and two objectively less similar domains (health, recreation).

2 Method

2.1 Data Set

The data analyses are based on all waves of the German Socio-Economic Panel (SOEP) from 1984 to 2005 (Wagner et al. 2006; Schimmack et al. 2008). The SOEP is an ongoing panel study based on a nationally representative household sample. Data were collected by annual interviews by a professional high-quality fieldwork organization (Infratest Social Research, Munich) using a variety of method, including interviews, mailed questionnaires, and more recently computer assisted personal interviews (CAPI). The data analyses are based on the international data set that is being made available to scientists who sign a contract with the DIW (German Institute for Economic Research). Further information about the sample, related publications, or forms to request the data used in this publication can be found on the SOEP website of the DIW (http://www.diw.de/deutsch/soep/29012.html).

For the present study, the sample is limited to participants who were married in 1984 and who completed all 22 waves from 1984 to 2005. The sample sizes varies from $N = 604$ to $N = 740$ due to the different number of missing data for different items. The dependent variables were satisfaction ratings on an 11-point scale ranging from $0 =$ completely dissatisfied to $10 =$ completely satisfied. At the beginning of the interview, respondents answered questions about domain satisfaction. Only domains that were included on all 22 waves were used for this article. Two domains were objectively identical (household income, housing) and two domains were not objectively identical (recreation, health). At the end of the interview, respondents answered a single question about satisfaction with life in general. The order of domain and global life-satisfaction questions may raise concerns that the questions about domain satisfaction primed information that was used when respondents judged life-satisfaction. However, these concerns are unfounded for two reasons. First, item-order effects have been examined in numerous studies and tend to be quite small (Schimmack and Oishi 2005). Second, the global item was only answered after numerous questions (often more than 100) about other aspects of people's lives. Thus, it is unlikely that respondents explicitly recalled answers to the domain satisfaction questions when they answered the life-satisfaction question.

Previous publications using the life-satisfaction question in the SOEP have demonstrated that it is reliable and valid (e.g., Lucas et al. 2004; Schimmack 2009; Schimmack et al. 2008), and that the item produces similar results to other longitudinal studies (Lucas and Donnellan 2007). A reliability study indicated that this item has an approximate 6-week retest correlation of $r = .5$ for new respondents, which increases to approximately

$r = .7$ for long-term participants in the panel study (Schimmack et al. 2008). These reliability estimates are comparable to those in other studies (Schimmack and Oishi 2005).

3 Results

Overall model fit was evaluated using standard values of a variety of fit indices (Schermelleh-Engel et al. 2003). Acceptable fit requires a comparative fit index (CFI) greater than .95, a root mean square error of approximation (RMSEA) smaller than .06, and a standard root mean residual (SRMR) smaller than .08. A variety of models were fitted to the data to test the stationarity assumptions in various ways (it is impossible to test all assumptions in a single model). All models produced adequate fit and similar results regarding the theoretically relevant parameters. Model fit could be further improved by relaxing the stationarity assumption for the reliability estimates. This finding is consistent with previous studies that show increasing reliability in panel data (Ehrhardt et al. 2000). However, substantial parameters of the model were not affected. To simplify the presentation of the results, only the results of a model with separate estimates of parameter estimates for the first and second half of the study are reported. This model was chosen because it shows the stability of parameter estimates over time.

We also fitted a mixed model to a data set with a larger number of married couples ($N = 10{,}115$). A mixed model allows to include couples who participated only one some waves. The parameter estimates for this model were very similar to those presented in this article. One potential problem of mixed models is that estimates of long-term stability are based on participants who remained in the study for a long time, whereas short-term stability estimates are more strongly influenced by couples who entered the panel recently or dropped out of the study. In this way, a mixed model mixes information from different units of analysis, which may produce misleading results. For this reason, we prefer to present the parameter estimates that are obtained based on a well-defined sample.

3.1 Life Satisfaction

Table 1 shows the standardized parameter estimates for this model. Rather than reporting error variance, the amount of reliable variance (1 − error variance) is reported. Furthermore, trait variance and state variance are reported as the proportion of the reliable variance rather than the total variance. The amount of new variance is not reported, but can be inferred from state stability (i.e., new variance = 1 − stability^2). For example, a stability of .9 implies that 81% of the variance is old and 19% of the variance is new. Table 1 also provides information about the differences between parameter estimates for the first and second half, in terms of the effect size measure q (Cohen 1988). This measure is based on Fisher r-z transformations of correlations and computing the difference between the two z-values. For comparisons of variances, the square root of the variance was used to compute effect sizes. The reason for this transformation is that the square root reveals the path coefficients for the assumed causal effects. For example, the square root of reliability coefficients reveals the causal effect of reliable variance on observed variance. Table 1 also reports the boundaries of the 99% confidence interval. If two confidence intervals are not overlapping, two parameters can be considered significantly different from each other.

Table 1 Fit indices for domain satisfaction

Domain	*N*	Chi	*df*	CFI	RMSEA	SRMR
Life satisfaction	725	1,710	967	.963	.033	.076
Household income	680	1,890	967	.959	.037	.069
Housing	607	1,899	967	.954	.040	.070
Recreation	693	1,691	967	.958	.033	.057
Health	740	1,377	967	.979	.024	.042

The comparison of the parameter estimates for the first and second half shows fairly similar estimates, and effect sizes of differences between parameter estimates were relatively small ($q < .40$). Especially, estimates of spousal similarity were quite similar across the two halves. Inspection of the theoretically important parameter estimates leads to the following conclusions.

1. The reliability of the single-item life satisfaction item ranges from .5 to .7, and tends to increase over time.
2. Trait and state variance contribute about equally to the reliable variance, although state variance tends to make a slightly stronger contribution.
3. The annual stability of state variance of life satisfaction is high, suggesting that individual differences in life satisfaction change only slowly over time. Furthermore, state stability may increase over time.
4. There is high ($r > .5$) spousal similarity in trait, initial state, new state, and final state variance, but relatively small spousal similarity in the error component.
5. Finally, it is noteworthy that parameter estimates for wives and husbands were very similar. This finding suggests that similar factors influence husbands' and wives' life satisfaction.

The model parameters imply that most of the state variance at the end of the study is cumulative new state variance, because the amount of shared variance is an exponential function of the time interval. Over the long period of 22 years, initial state variance contributes only 2% of the final state variance, despite high annual stability ($[.91^{22}]^2 = .02$). As a result, about 50% of the reliable variance in life-satisfaction can be attributed to factors that change over time. Another important finding was the high spousal similarity in trait variance.

3.2 Domain Satisfaction

Applying the same model to the domain satisfaction data also produced acceptable fit (Table 1). The parameter estimates for the four domains are shown in Tables 2, 3, 4 and 5. As for life-satisfaction, parameter estimates tend to be quite similar for the first and second half of the study as well as for wives and husbands. Ratings of domain satisfaction also show increasing reliability over time.

The most important finding is that estimates of spousal similarity in state variance vary across domains in a theoretically predicted manner. Domains that are objectively identical for both spouses showed very high spousal similarities with estimates ranging from .73 to .83. In contrast, domains that are objectively less similar produced less spousal similarity. The lowest similarity was obtained for health satisfaction, .22–.26. A difference between a

Table 2 Parameter estimates for life satisfaction

Parameter	1–11	12–22	*Q*
Wives			
Reliability	.58 [.54l.62]	.64 [.61l.67]	.10
Trait variance	.50 [.37l.63]	.39 [.27l.51]	−.15
State variance	.50 [.37l.63]	.61 [.49l.73]	.17
Annual stability	.87 [.83l.92]	.92 [.89l.94]	.26
Husbands			
Reliability	.57 [.54l.61]	.67 [.64l.70]	.17
Trait variance	.43 [.30l.56]	.38 [.25l.50]	−.07
State variance	.57 [.44l.70]	.62 [.50l.75]	.08
Annual stability	.88 [.84l.92]	.92 [.89l.94]	.21
Similarity			
Trait	.77 [.69l.86]		
Initial state	.74 [.61l.87]	.62 [.53l.71]	−.23
New state	.61 [.52l.70]	.63 [.55l.72]	.03
Final state	.64 [.56l.73]	.63 [.57l.69]	−.02
Error	.32 [.28l.36]	.25 [.21l.29]	−.09

Table 3 Parameter estimates for household income

Parameter	1–11	12–22	*Q*
Wives			
Reliability	.62 [.59l.66]	.70 [.67l.73]	.15
Trait variance	.35 [.23l.48]	.49 [.37l.61]	.19
State variance	.65 [.52l.77]	.51 [.39l.63]	−.22
Annual stability	.87 [.83l.91]	.93 [.90l.95]	.33
Husbands			
Reliability	.60 [.63l.67]	.68 [.65l.71]	.14
Trait variance	.43 [.30l.56]	.45 [.34l.57]	.03
State variance	.57 [.44l.70]	.55 [.43l.66]	−.03
Annual stability	.87 [.83l.90]	.91 [.88l.94]	.19
Similarity			
Trait	.93 [.88l.98]		–
Initial state	.81 [.73l.88]	.80 [.73l.86]	−.03
New state	.82 [.77l.88]	.79 [.73l.86]	−.09
Final state	.82 [.77l.87]	.79 [.75l.84]	−.09
Error	.33 [.29l.37]	.28 [.24l.32]	−.06

.82 and a .22 correlation is very large, $q = 1.87$ (Cohen 1988). Thus, the results provide strong support for the hypothesis that spousal similarity in the state variance of domain satisfaction reflects the influence of environmental factors on domain satisfaction. Another important finding is that spouses are also similar in the trait component of domain satisfaction, and that trait similarity often exceeded state similarity. This finding cannot be attributed to differences in reliabilities because both estimates are based on latent factors that control for measurement error (Table 6).

Table 4 Parameter estimates for housing

Parameter	1–11	12–22	Q
Wives			
Reliability	.66 [.62|.70]	.67 [.63|.70]	.02
Trait variance	.34 [.23|.46]	.34 [.21|.47]	.00
State variance	.66 [.54|.77]	.66 [.53|.79]	.00
Annual stability	.83 [.79|.88]	.91 [.88|.94]	.34
Husbands			
Reliability	.67 [.63|.70]	.67 [.63|.70]	.00
Trait variance	.43 [.31|.56]	.43 [.31|.56]	.00
State variance	.57 [.44|.69]	.57 [.43|.66]	.00
Annual stability	.82 [.77|.87]	.91 [.88|.94]	.37
Similarity			
Trait	.89 [.84|.96]		–
Initial state	.86 [.79|.93]	.83 [.77|.89]	−.11
New state	.83 [.79|.88]	.73 [.65|.81]	−.26
Final state	.83 [.79|.88]	.74 [.69|.79]	−.24
Error	.40 [.35|.44]	.22 [.18|.26]	−.20

Table 5 Parameter estimates for recreation

Parameter	1–11	12–22	Q
Wives			
Reliability	.57 [.53|.61]	.65 [.62|.69]	.13
Trait variance	.32 [.15|.48]	.31 [.16|.47]	−.01
State variance	.68 [.52|.85]	.69 [.53|.84]	.02
Annual stability	.86 [.81|.92]	.93 [.91|.95]	.37
Husbands			
Reliability	.54 [.51|.57]	.62 [.58|.65]	.13
Trait variance	.20 [.02|.38]	.27 [.07|.47]	.09
State variance	.80 [.62|.98]	.73 [.53|.93]	−.17
Annual stability	.90 [.87|.94]	.93 [.90|.96]	.19
Similarity			
Trait	.89 [.62|.1.00]		–
Initial state	.43 [.28|.58]	.45 [.35|.55]	.02
New state	.49 [.39|.59]	.46 [.35|.57]	−.04
Final state	.48 [.38|.57]	.46 [.39|.54]	−.03
Error	.20 [.16|.24]	.18 [.15|.22]	−.02

4 Discussion

The main contribution of this article was to provide strong empirical evidence for environmental influences on life satisfaction and domain satisfaction because spouses' life satisfaction and domain satisfaction changed in the same direction. Furthermore, the amount of spousal similarity varied as a function of the domain. Spousal similarity was higher for objectively identical domains than for objectively less similar domains. As a

Table 6 Parameter estimates for health

Parameter	1–11	12–22	*Q*
Wives			
Reliability	.58 [.55|.62]	.68 [.64|.71]	.17
Trait variance	.59 [.49|.65]	.63 [.56|.70]	.07
State variance	.41 [.35|.51]	.37 [.30|.44]	−.05
Annual stability	.83 [.78|.89]	.83 [.76|.89]	.00
Husbands			
Reliability	.62 [.59|.70]	.66 [.63|.69]	.07
Trait variance	.46 [.37|.56]	.54 [.45|.63]	.11
State variance	.54 [.44|.63]	.46 [.37|.55]	−.11
Annual stability	.85 [.81|.89]	.87 [.82|.92]	.08
Similarity			
Trait	.47 [.36|.58]		–
Initial state	.24 [.06|.42]	.11 [− .05|.27]	−.13
New state	.26 [.13|.39]	.22 [.09|.35]	−.04
Final state	.26 [.13|.38]	.22 [.12|.32]	−.04
Error	.14 [.10|.18]	.08 [.04|.13]	−.06

result, SWB can provide meaningful information about the quality of individuals' lives as assessed from the subjective perspective of the individuals living these lives, and these judgments can inform public policies (Diener et al. 2009; Veenhoven 1994). Beyond these broader implications, the present study also raises important questions about the interpretation of previous studies of environmental versus genetic influences on SWB. These implications are discussed separately for the findings regarding state similarity and trait similarity in life satisfaction.

4.1 Spousal Similarity in the State Component of Life Satisfaction

Behavioral genetics studies typically find that genetic factors explain 25–50% of the variance in various measures of well-being (Lykken and Tellegen 1996; Nes et al. 2006; Stubbe et al. 2005; Weiss et al. 2008). Furthermore, the studies suggest that the shared environment with a co-twin has no influence on SWB. As a result, the remaining variance is either due to environmental influences unique to each twin or measurement error. These findings seem to generalize across different well-being measures, including measures of life satisfaction (Stubbe et al. 2005; Weiss et al. 2008). Furthermore, even reliable environmental influences may be due to short-lived temporary changes. The uncertainty about the nature of the variance that is not explained by genetic factors has led some authors to suggest that these factors are relatively unimportant for the understanding of individual differences in SWB, and have created a perception in the field that well-being is stable and genetically predetermined (cf. Diener et al. 2006).

The present findings show that non-heritable variance is far from irrelevant. Consistent with several other studies, this study shows that state variance in happiness changes only gradually over time (Ehrhardt et al. 2000; Fujita and Diener 2005; Lucas and Donnellan 2007; Schimmack and Oishi 2005; Veenhoven 1994). More importantly, this study shows that environmental factors, directly or in interaction with genetic dispositions, cause some of the changes in life satisfaction and satisfaction with some life domains. Although the exact environmental factors remain unclear, it is plausible that some of these environmental factors are under people's control, suggesting that it is possible for individuals to increase their well-being. Moreover, genetic factors may not be the only reason for stability in well being. Environmental barriers (e.g., few jobs that pay $1,000/h, few six-bedroom houses in Manhattan, and the price of a Porsche) may also contribute to the difficulty of increasing well-being.

Another reason for divergent conclusions about genetic versus environmental influences could be the focus on different aspects of well-being. The present study examined cognitive well-being and found clear evidence for environmental influences on life-satisfaction and domain satisfaction. In contrast, Lykken and Tellegen (1996) used the "well-being" subscale of the Multidimensional Personality Questionnaire (Tellegen 1982). This scale is intended to measure an individual's "*disposition* [italics added] to feel good about oneself and one's own corner of the world" (Lykken and Tellegen 1996, p. 186). By definition, individual's dispositions should be less sensitive to environmental influences. At the same time, it is questionable to assess well-being with a dispositional measure. Even individuals with a happy disposition can have low well-being in adverse circumstances. Consistent with this idea, Schimmack et al. (2008) found that unemployment was strongly (negatively) related to life-satisfaction, but unrelated to measures of personality dispositions. This finding suggests that life-events such as unemployment influence well-being, but not personality dispositions that influence well-being.

One limitation of the present work is that the present findings do not show the effect size of environmental effects on SWB. The reason is that various causal models can explain spousal similarity in SWB. One causal model assumes that shared environment factors have independent and equal effects on both spouses. Based on this model, spousal similarity would be equivalent to the amount of variance that is explained by shared environmental factors. For life satisfaction, this would imply that 60% of the state variance is explained by the shared environment between spouses. The remaining variance could be due to unshared environmental factors or even genetic factors, if genetic factors also contribute to changes in life satisfaction (Nes et al. 2006).

However, an alternative causal model could assume that spousal similarity is only partially an environmental effect. According to this model, husbands' genes have a strong influence on husbands' SWB, but the same genes also influence wives' SWB. For example, a husband may have a genetic disposition to be sensitive to stress. During times of stress this disposition leads to a marked decrease in the husband's life satisfaction, which in turn influences his wife's life satisfaction. A similar process could produce an influence of wives' genetic dispositions on husbands' life satisfaction. This model can predict high spousal similarity with relatively weak environmental effects. However, two findings suggest that spousal similarity in state variance is caused by shared environment effects. First, the higher spousal similarity in domain satisfaction in objectively identical domains than in objectively less similar domains is more consistent with a shared environment effect. Second, longitudinal behavioral genetics studies suggest that genes make a small contribution to state variance (Lykken and Tellegen 1996; Nes et al. 2006). Nevertheless, it is important to note that estimates of effect sizes depend on assumptions about the causal model that produces spousal similarity. An important avenue for future research is to test these assumptions.

4.2 Spousal Similarity in the Trait Component of Life Satisfaction

We found high spousal similarity in the trait component of life satisfaction and domain satisfaction. Schimmack et al. (2008) demonstrated that the same trait underlies global and domain satisfaction judgments. Presumably, this trait reflects a general disposition to be more or less easily satisfied with life circumstances (Schimmack 2009). The finding of high spousal similarity in this disposition has important implications for the interpretation of twin studies of well-being. The reason is that twin studies implicitly assumed that spouses do not have similar genetic dispositions that influence well-being to set genetic similarity of dizygotic twins at 50%. Our findings challenge this assumption because twin studies typically assume that most of the trait variance is heritable. However, spousal similarity in the trait component violates either the assumption that spouses do not share genetic dispositions or the assumption that the trait component is mostly heritable. The present data are insufficient to reveal which of the assumptions is invalid, but it is instructive to examine the implications of spousal similarity in genetic dispositions that influence well-being.

A model of phenotypic assortative mating assumes that spouses choose each other on the basis of similarity in phenotypes that are influenced by similar genetic dispositions. Based on this model, it is possible to infer genetic similarity between spouses from the observed similarity between spouses with a simple formula (Neale and Cardon 1992):

$$\text{Genetic similarity} = \text{Phenotypic similarity} \times \text{Heritability}\ (h^2)$$

Previous studies have suggested that 80% of the trait variance in SWB is heritable (Lykken and Tellegen 1996; Nes et al. 2006). As a result, the phenotypic trait similarity of about $r = .8$ in the present study implies a genetic similarity of 64% for the trait component of life-satisfaction (i.e., .8 × .8). As the trait component accounts for about 50% of the reliable total variance, this estimate implies a genetic similarity of 32% for the total reliable variance. This estimate can be entered into another formula to obtain an estimate of genetic similarity for DZ twins (Reynolds et al. 1996):

$$\text{DZ genetic similarity} = 0.5 \times (1 + \text{Parents' genetic similarity}).$$

According to this formula, genetic similarity for the trait component increases from the standard estimate of 50% under the assumption of random mating to 66% (0.5 × 1.32). The implications of this new estimate of genetic similarity between DZ twins for twin studies can be illustrated with actual DZ twin or sibling correlations in the literature. For example, two recent studies of life-satisfaction have reported DZ correlations of about $r = .10$ (Stubbe et al. 2005; Weiss et al. 2008). Much higher MZ correlations suggest that shared (rearing) environment does not contribute to this correlation. A DZ correlation of .10 would is typically interpreted as evidence that 20% of the variance is due to additive genetic effects (Weiss et al. 2008), based on the assumption that DZ twins share 50% of their genes (.20 = .10/.50). However, the estimated genetic similarity in this study suggests that additive genetic effects make a smaller contribution (.15 = .10/.66) because the genetic similarity of DZ twins is higher. The same implications apply for parent–child similarity, which also tends to be quite low (Casas et al. 2008).

Our findings also have important implications for the modeling of additive and non-additive genetic effects. It is typically assumed that genetic effects are additive when MZ correlations are approximately twice the size of DZ correlations. Genetic effects are assumed to be non-additive if MZ correlations exceed DZ correlations times 2. However, the factor 2 in this formula is based on the assumption that genetic similarity between DZ twins is 50%. If genetic similarity is higher (e.g., 66%), an additive model predicts that MZ correlations should match DZ correlations multiplied by a smaller factor (1.52). Using this criterion, even the data of studies that have assumed additive genetic effects are inconsistent with additive genetic models (Weiss et al. 2008). Non-additive genetic effects create several problems for genetic models of well-being. First, it will be difficult to find genetic markers that predict well-being measures because the effects of one gene are moderated by the effects of other genes. Second, it casts doubt on the contribution of personality traits to well-being because personality data tend to be consistent with additive genetic models. However, if personality traits are influenced by additive genetic models and well-being is influenced by non-additive genetic factors, it suggests that personality traits alone are insufficient to account for the heritability of well-being. In short, another contribution of this study is to raise concerns about the standard assumption in twin studies that spouses do not have similar genetic dispositions that influence well-being. This issue requires more attention in behavioral genetics studies of well-being.

The distinction between additive and non-additive genetic effects is not merely an academic question. It also has practical implications. For additive genetic phenotypes, parents can expect similar phenotypes in their children. For example, tall parents can expect to have tall children, because height is largely determined by additive genetic effects. In contrast, if well-being is largely a non-additive phenotype, parents with a happy disposition cannot expect their children to inherit the same happy disposition. This may have been the reason why Lykken and Tellegen (1996) referred to happiness as a "stochastic phenomenon." Even though a large portion of the stable component of well-being may be influenced by genetic dispositions, parents' phenotype is an unreliable predictor of children's disposition for high well-being.

4.3 Limitations

The main limitation of DLPA is that it can only reveal environmental factors that are shared between members of a dyad. In this regard, DLPA is overly conservative and is likely to underestimate environmental effects. Thus, it is important to compare the results of DLPA with results from other studies, most prominently twin studies. Twin studies overestimate environmental effects whenever measurement error cannot be estimated. Thus, twin similarity and spousal similarity provide a range for plausible estimates of environmental effects.

A practical limitation is that DLPA requires dyadic data with repeated assessments over a sufficiently long interval to observe changes in state variance. Thus, DLPA cannot be applied to most data sets in the well-being literature that are often limited to two or three assessments over relatively short time intervals (Schimmack and Oishi 2005). This may explain why this method has not been used more often despite its obvious advantages to test causal models. Well being scientists need to initiate more longitudinal studies to take advantage of DLPA and other innovations in the modeling of longitudinal data. By demonstrating the power of DLPA to demonstrate environmental influences, we hope that more studies with adequate data will become available. For now, DLPA can be applied to existing data sets such as the SOEP and other household panel studies.

Another limitation of DLPA is that it does not reveal the nature of environmental factors. In this regard, DLPA is akin to traditional behavioral genetics studies. Behavioral genetics studies demonstrated that genes are important, but could not reveal which genes are important. Currently, molecular genetics studies are starting to provide answers to this question. Similarly, DLPA only demonstrates that shared environmental factors influence well-being, but it does not reveal the nature of these factors. To examine the nature of these factors, it is necessary to include relevant predictor variables in the longitudinal study. For example, Lucas et al. (2004) demonstrated that unemployment has negative effects on individuals' life satisfaction. This environmental effect on individuals' life satisfaction may contribute to dyadic similarity if unemployment of one spouse also causes changes in the other spouse's life satisfaction. To test this hypothesis, additional predictor variables can be included in a DLPA to examine whether they contribute to spousal similarity.

Another limitation of DLPA is the requirement to make assumptions about stationarity. However, this assumption is not necessary and can be tested. We tested this assumption using a split-waves design and found no major violations of this assumption.

Another potential limitation of the present study is the assessment of life-satisfaction and domain satisfaction with single items. One common belief is that single-item measures are undesirable because it is impossible to determine their reliability. However, the present article demonstrates that this is not a major concern in longitudinal panel studies. In longitudinal panel studies it is possible to estimate reliability of single-item measures on the basis of the pattern of retest correlations (Heise 1969; Kenny and Zautra 1995). The reliability estimates of the single-item measures in this study are consistent with previous estimates in the literature (Schimmack and Oishi 2005). The main limitation of single-item indicators is that single-item measures are bound to have lower validity than multiple-item scales because the specific wording of the single item is likely to produce some systematic error variance that can be minimized by averaging across multiple items. However, validity coefficients for single-item measures tend to be very high (Saris et al. 1998). Thus, it is likely that multiple-item measures of the same construct would produce very similar results.

Arguably, a much bigger problem is that self-report measures may be only partially valid measures of well-being. Self-informant correlations for well-being are moderate (Schneider and Schimmack 2009). Thus, an important area for future research is the assessment of well-being with reports by multiple informants (Lucas et al. 1996; Schimmack and Diener 2003; Walker and Schimmack 2008).

In conclusion, the main novel contribution of this article is the demonstration that spouses well-being changes in the same direction over time. This finding provides strong evidence for environmental influences on well-being. Future research needs to examine, the nature of these environmental factors and how they interact with personality variables to produce well-being.

Acknowledgments We would like to thank Jennifer Tackett, Kelly L. Klump, Lindon Eaves, and Shigehiro Oishi, and Ivana Anusic for helpful comments. The preparation of this article was supported by a standard research grant of the Canadian Social Sciences and Humanities Research Council awarded to the first author.

References

Alley, T. R. (1993). The developmental stability of facial attractiveness: New longitudinal data and a review. *Merrill-Palmer Quarterly, 39*(2), 265–278.

Bookwala, J., & Schulz, R. (1996). Spousal similarity in subjective well-being: The cardiovascular health study. *Psychology and Aging, 11*(4), 582–590.

Casas, F., Coenders, G., Cummins, R. A., Gonzalez, M., Figuer, C., & Malo, S. (2008). Does subjective well-being show a relationship between parents and their children? *Journal of Happiness Studies, 9*(2), 197–205.

Caspi, A., Herbener, E., & Ozer, D. (1992). Shared experiences and the similarity of personalities: A longitudinal study of married couples. *Journal of Personality and Social Psychology, 62*(2), 281–291.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cole, D. A., Martin, N. C., & Steiger, J. H. (2005). Empirical and conceptual problems with longitudinal trait-state models: Introducing a trait-state-occasion model. *Psychological Methods, 10*(1), 3–20.
Diener, E. (1984). Subjective well-being. *Psychological Bulletin, 95*(3), 542–575.
Diener, E., Lucas, R. E., Schimmack, U., & Helliwell, J. (2009). *Well-being for public policy*. New York: Oxford University Press.
Diener, E., Lucas, R. E., & Scollon, C. N. (2006). Beyond the hedonic treadmill: Revising the adaptation theory of well-being. *American Psychologist, 61*(4), 305–314.
Eaves, L. J., Fulker, D. W., & Heath, A. C. (1989). The effects of social homogamy and cultural inheritance on the covariances of twins and their parents: A LISREL model. *Behavior Genetics. Special Issue: Twin methodology using LISREL, 19*(1), 113–122.
Ehrhardt, J. J., Saris, W. E., & Veenhoven, R. (2000). Stability of life-satisfaction over time: Analysis of change in ranks in a national population. *Journal of Happiness Studies, 1*(2), 177–205.
Feingold, A. (1988). Matching for attractiveness in romantic partners and same-sex friends: A meta-analysis and theoretical critique. *Psychological Bulletin, 104*(2), 226–235.
Fujita, F., & Diener, E. (2005). Life satisfaction set point: Stability and change. *Journal of Personality and Social Psychology, 88*(1), 158–164.
Headey, B., Veenhoven, R., & Wearing, A. (1991). Top-down versus bottom-up theories of subjective well-being. *Social Indicators Research, 24*(1), 81–100.
Heath, A. C., & Eaves, L. J. (1985). Resolving the effects of phenotype and social background on mate selection. *Behavior Genetics, 15*(1), 15–30.
Heise, D. R. (1969). Separating reliability and stability in test–retest correlation. *American Sociological Review, 34*(1), 93–101.
Kenny, D. A. (1975). Cross-lagged panel correlation: A test for spuriousness. *Psychological Bulletin, 82*(6), 887–903.
Kenny, D. A., & Zautra, A. (1995). The trait state error model for multiwave data. *Journal of Consulting and Clinical Psychology, 63*(1), 52–59.
Lucas, R. E., Clark, A. E., Georgellis, Y., & Diener, E. (2004). Unemployment alters the set point for life satisfaction. *Psychological Science, 15*(1), 8–13.
Lucas, R. E., Diener, E., & Suh, E. (1996). Discriminant validity of well-being measures. *Journal of Personality and Social Psychology, 71*(3), 616–628.
Lucas, R. E., & Donnellan, M. B. (2007). How stable is happiness? Using the STARTS model to estimate the stability of life satisfaction. *Journal of Research in Personality, 41*(5), 1091–1098.
Lykken, D., & Tellegen, A. (1996). Happiness is a stochastic phenomenon. *Psychological Science, 7*(3), 186–189.
McCartney, K., Harris, M. J., & Bernieri, F. (1990). Growing up and growing apart: A developmental meta-analysis of twin studies. *Psychological Bulletin, 107*(2), 226–237.
Mischel, W. (1968). *Personality and assessment*. New York: Wiley.
Muthén, L. K., & Muthén, B. O. (2007). *Mplus user's guide* (Fourth Edition ed.). Los Angeles, CA: Muthén & Muthén.
Neale, M. C., & Cardon, L. R. (1992). Methodology for genetic studies of twins and families. *NATO ASI Series D: Behavioural and Social Sciences, 67*(496).
Nes, R. B., Roysamb, E., Tambs, K., Harris, J. R., & Reichborn-Kjennerud, T. (2006). Subjective well-being: Genetic and environmental contributions to stability and change. *Psychological Medicine, 36*(7), 1033–1042.
Ormel, J., & Rijsdijk, F. V. (2000). Continuing change in neuroticism during adulthood: Structural modelling of a 16-year, 5-wave community study. *Personality and Individual Differences, 28*(3), 461–478.
Ormel, J., & Schaufeli, W. B. (1991). Stability and change in psychological distress and their relationship with self-esteem and locus of control: A dynamic equilibrium model. *Journal of Personality and Social Psychology, 60*(2), 288–299.
Reynolds, C. A., Baker, L. A., & Pedersen, N. L. (1996). Models of spouse similarity: Applications to fluid ability measured in twins and their spouses. *Behavior Genetics, 26*(2), 73–88.
Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology. McGraw Hill series in social psychology* (p. 286). New York, NY, UK: Mcgraw Hill Book Company.
Roysamb, E., Harris, J. R., Magnus, P., Vitterso, J., & Tambs, K. (2002). Subjective well-being: Sex-specific effects of genetic and environmental factors. *Personality and Individual Differences, 32*(2), 211–223.
Rudinger, G., & Rietz, C. (1998). The neglected time dimension? Introducing a real longitudinal model: Testing latent growth curves, stability, and reliability as time-bound processes. *Methods of Psychological Research. Special Issue: Analysis of longitudinal data, 3*(2), 109–130.
Saris, W. E., Van Wijk, T., & Scherpenzeel, A. (1998). Validity and reliability of subjective social indicators—the effect of different measures of association. *Social Indicators Research, 45*(1–3), 173–199.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research, 8*(2), 23–74.

Schilling, O. (2006). Development of life satisfaction in old age: Another view on the "Paradox". *Social Indicators Research, 75*(2), 241–271.

Schimmack, U. (2009). Well-being: Measuring wellbeing in the SOEP. *Schmollers Jahrbuch, 129*, 1–9.

Schimmack, U., & Diener, E. (2003). Predictive validity of explicit and implicit self-esteem for subjective well being. *Journal of Research in Personality, 37*(2), 100–106.

Schimmack, U., & Lucas, R. E. (2007). Marriage matters: Spousal similarity in life satisfaction. *Journal of Applied Social Science Studies, 127*, 1–7.

Schimmack, U., & Oishi, S. (2005). The influence of chronically and temporarily accessible information on life satisfaction judgments. *Journal of Personality and Social Psychology, 89*(3), 395–406.

Schimmack, U., Schupp, J., & Wagner, G. G. (2008). The influence of environment and personality on the affective and cognitive component of subjective well-being. *Social Indicators Research, 89*, 41–60.

Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods, 1*(2), 199–223.

Schneider, L., & Schimmack, U. (2009). Self-informant agreement in well-being ratings: A meta-analysis. *Social Indicators Research*. doi:10.1007/s11205-009-9440-y.

Silventoinen, K., Kaprio, J., Lahelma, E., Viken, R. J., & Rose, R. J. (2003). Assortative mating by body height and BMI: Finnish twins and their spouses. *American Journal of Human Biology, 15*(5), 620–627.

Spotts, E. L., Neiderhiser, J. M., Towers, H., Hansson, K., Lichtenstein, P., Cederblad, M., et al. (2004). Genetic and environmental influences on marital relationships. *Journal of Family Psychology, 18*(1), 107–119.

Steyer, R., Schmitt, M., & Eid, M. (1999). Latent state-trait theory and research in personality and individual differences. *European Journal of Personality, 13*, 389–408.

Stubbe, J. H., Posthuma, D., Boomsma, D. I., & De Geus, E. J. C. (2005). Heritability of life satisfaction in adults: A twin-family study. *Psychological Medicine, 35*(11), 1581–1588.

Sumner, L. W. (1996). *Welfare, happiness, and ethics*. Oxford: Claredon Press.

Tambs, K., & Moum, T. (1992). No large convergence during marriage for health, lifestyle, and personality in a large sample of Norwegian spouses. *Journal of Marriage & the Family, 54*(4), 957–971.

Tellegen, A. (1982). *Brief manual for the multidimensional personality questionnaire*. Unpublished manuscript, University of Minnesota, Minneapolis.

Tesser, A. (1993). The importance of heritability in psychological research: The case of attitudes. *Psychological Review, 100*(1), 129–142.

Tower, R. B., & Kasl, S. V. (1996). Depressive symptoms across older spouses: Longitudinal influences. *Psychology and Aging, 11*(4), 683–697.

Veenhoven, R. (1994). Is happiness a trait? Tests of the theory that a better society does not make people any happier. *Social Indicators Research, 32*(2), 101–160.

Wagner, G. G., Frick, J. R., & Schupp, J. (2006). *Enhancing the power of household panel studies—the case of the German Socio-Economic Panel Study (SOEP)*. DIW data documentation no. 13.

Walker, S. S., & Schimmack, U. (2008). Validity of a happiness implicit association test as a measure of subjective well being. *Journal of Research in Personality, 42*(2), 490–497.

Weiss, A., Bates, T. C., & Luciano, M. (2008). Happiness is a personal(ity) thing—The genetics of personality and well-being in a representative sample. *Psychological Science, 19*(3), 205–210.

Zebrowitz, L. A., Olson, K., & Hoffman, K. (1993). Stability of babyfaceness and attractiveness across the life span. *Journal of Personality and Social Psychology, 64*(3), 453–466.