

# Day-To-Day Affect is Surprisingly Stable: A 2-Year Longitudinal Study of Well-Being

Nathan W. Hudson<sup>1</sup>, Richard E. Lucas<sup>1</sup>, and M. Brent Donnellan<sup>2</sup>

Social Psychological and  
Personality Science  
1-10

© The Author(s) 2016

Reprints and permission:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/1948550616662129

spps.sagepub.com



## Abstract

Previous research suggests global assessments of cognitive well-being—life satisfaction—are relatively stable over time. Far fewer studies have examined the extent to which experiential measures of affective well-being—the moods/emotions people regularly experience—are stable, especially over extended periods of time. The present study used longitudinal data from a representative sample of Germans to investigate the long-term stability of different components of well-being. Participants provided global ratings of life satisfaction and affect, along with experiential measures of well-being up to 3 times over 2 years. Results indicated between one-third and one half of the variance in people's daily affect was attributable to trait-like latent variables. Replicating meta-analytic findings, 50% of the variance in global measures of well-being was attributable to trait-like latent variables.

## Keywords

well-being, personality processes, day reconstruction method

Subjective well-being is a broad construct that reflects people's overall appraisals of the positivity of their lives, as well as the balance of their affective states (Diener, 1984). Attaining and sustaining well-being is deeply valued across the echelons of society—from individual persons (Diener & Oishi, 2004; Lucas & Diener, 2008) to national governments (Samuel, 2009; Stratton, 2010; University of Waterloo, 2011; U.S. Department of Health and Human Services, 2014). One critical question, therefore, is the degree to which well-being is driven by transitory situational forces (e.g., that may be responsive to interventions) versus stable individual differences.

Quantifying the degree to which well-being is driven by transitory versus enduring influences is complicated because well-being has multiple components, which vary with respect to at least two factors (Diener, 1984; Lucas & Diener, 2008; Lucas, Diener, & Suh, 1996). First, well-being includes both *cognitive* and *affective* aspects. Cognitive well-being refers to the extent to which individuals appraise their lives positively. In contrast, affective well-being refers to the moods and emotions people actually experience. Importantly, although cognitive and affective well-being are related, they are separable (for a review, see Busseri & Sadava, 2011). It is possible, for instance, that someone might experience frequent negative affect, while still believing his or her life is satisfying.

In addition to the distinction between cognitive and affective well-being, these components can be assessed using both *global* and *experiential* measures. Global measures (sometimes called *evaluative* measures) capture appraisals of overall life satisfaction or patterns of affect, whereas experiential measures assess lived experiences and momentary reports of well-being.

As depicted in Figure 1, the cognitive/affective and global/experiential factors cross to form four subtypes of well-being measures. Specifically, *global cognitive* measures capture individuals' beliefs about the general positivity of their lives (e.g., life satisfaction), whereas *experiential cognitive* measures tap people's in vivo appraisals of their circumstances (e.g., participants' satisfaction with momentary experiences). Similarly, *global affective* measures assess people's beliefs about their typical patterns of positive and negative moods/emotions, whereas *experiential affective* measures capture the extent to which individuals report actually experiencing positive and negative moods/feelings.

Most well-being research has focused on global cognitive measures (e.g., Schimmack & Oishi, 2005). However, the various subtypes of well-being are separable (e.g., Kim-Prieto, Diener, Tamir, Scollon, & Diener, 2005; Lucas et al., 1996). For instance, participants' global evaluations of how frequently they experience various emotions correlate only moderately with measures of their actual felt-affect ( $r \sim .20-.30$ ; Anusic, Lucas, & Donnellan, 2016b). Researchers are divided with respect to whether this discrepancy indicates global measures are less valid than experiential ones (e.g., Robinson & Clore,

<sup>1</sup> Michigan State University, East Lansing, MI, USA

<sup>2</sup> Texas A&M University, College Station, TX, USA

## Corresponding Author:

Nathan W. Hudson, Department of Psychology, Michigan State University, 316 Physics Road, East Lansing, MI 48824, USA.

Email: nwhudson@msu.edu.

	Global	Experiential
Cognitive	<p><b>Global/overall life satisfaction</b></p> <p>Satisfaction with Life Scale; Single-item life satisfaction</p>	<p><b>Appraisals of/satisfaction with situations</b></p> <p>Experience sampled/day reconstructed appraisals</p>
Affective	<p><b>Global affect; perceptions of one's patterns of affect</b></p> <p>Positive and Negative Affect Schedule</p>	<p><b>Experienced affect</b></p> <p>Experience sampled/day reconstructed felt affect</p>

**Figure 1.** Well-being can be subdivided by two factors (global vs. experiential; cognitive vs. affective) to produce four separate subtypes of well-being. Descriptions of each subcomponent of well-being, along with sample measures thereof, are listed in each cell.

2002a), or whether both types of measures simply tap different facets of well-being (e.g., Kim-Prieto et al., 2005). Irrespective of this debate, the various components of well-being may be differentially influenced by situational versus enduring forces. For example, higher income predicts modestly greater life satisfaction (e.g., Lucas & Schimmack, 2009) but not happiness (e.g., Hudson, Lucas, Donnellan, & Kushlev, 2016). Thus, fully understanding the extent to which different aspects of well-being are determined by fleeting circumstances versus enduring factors requires studying each component separately (Kim-Prieto et al., 2005; Tay, Chan, & Diener, 2014).

### To What Extent is Well-Being Influenced by Transient Versus Enduring Factors?

One approach to investigating the extent to which well-being is determined by situational versus trait-like forces is to identify and test theoretically relevant predictors. To this end, research has examined the degree to which specific factors—such as socioeconomic status (Pinquart & Sorensen, 2000), health (Okun, Stock, Haring, & Witter, 1984), disability (Lucas, 2007), relationships (Lucas & Dyrenforth, 2006), and personality (e.g., Dyrenforth, Kashy, Donnellan, & Lucas, 2010; Schimmack, Radhakrishnan, Oishi, Dzokoto, & Ahadi, 2002)—correlate with global well-being. These studies have generally found both malleable circumstances (e.g., income) and enduring individual differences (e.g., personality) have moderate associations with global well-being. Far fewer

studies, in contrast, have focused on the correlates of experiential well-being (cf. Anusic, Lucas, & Donnellan, 2016a; Anusic et al., 2016b).

A second, more omnibus approach to quantifying the extent to which well-being is driven by relatively constant versus fluid factors is to examine its stability over time. Specifically, to the extent that well-being is driven by situational forces—even relatively enduring, albeit ultimately impermanent ones—its test–retest stability will decay over increasingly long test–retest intervals (Fraley & Roberts, 2005; Fraley, Vicary, Brumbaugh, & Roisman, 2011). In contrast, to the extent that well-being is determined by unchanging forces, its test–retest stability will eventually plateau at some nonzero lower bound, even over indefinitely long periods of time (Fraley et al., 2011; Fraley & Roberts, 2005; Kenny & Zautra, 1995).

Thus, it is possible to infer the extent to which well-being is determined by transitory forces, versus permanent ones, by examining its test–retest stability over varying periods of time. To this end, meta-analyses suggest test–retest stabilities in global cognitive well-being are approximately  $r = .60$ ,  $.50$ , and  $.35$  over 2, 5, and 10 years, respectively—asymptotically approaching a lower bound of  $.20$ – $.35$  (Anusic & Schimmack, 2016; Schimmack & Oishi, 2005).

Far less research has examined the degree to which experiential well-being is stable over long periods of time, although one study suggests stability in experience-sampled positive affect is approximately  $r = .60$  over both 5- and 10-years (Carstensen et al., 2011).<sup>1</sup> It is therefore unclear whether people's patterns of felt affect are more or less stable than global well-being. This lack of knowledge is exacerbated by the fact that stability in experiential measures may vary—not only as a function of the length of the test–retest interval (Fraley et al., 2011; Fraley & Roberts, 2005)—but also as a function of the window of time across which experiences are aggregated. For example, research suggests people's moment-by-moment emotions are essentially random and do not correlate across time (Epstein, 1979); however, once aggregated, these emotional experiences increase in temporal stability (Anusic et al., 2016a; Diener & Larsen, 1984). To avoid burdening respondents, however, experiential measures are often assessed over only a single day—especially when used in the context of large-scale survey work (e.g., Anusic et al., 2016b). It is unknown whether such a short aggregation window is sufficient to capture stable variance in people's affective experiences, or whether aggregating over longer periods of time (e.g., weeks, months) would be required to capture reliable individual differences in experiential well-being.

### Overview of the Present Study

The present study was designed to compare the stability of experiential and global well-being. Up to 3 times over 2 years, participants' experiential well-being was measured using the day reconstruction method (DRM; Kahneman, Krueger, Schkade, Schwarz, & Stone, 2004). As an alternative to experience sampling methods (ESM; Shiffman, Stone, & Hufford,

2008), in the DRM, respondents categorize their prior day in terms of “episodes” and then rate their affective experiences during these episodes. In contrast to ESM, which requires specialized equipment and is intrusive for participants, DRM measures can be administered via standard survey format, and some versions can be completed in as few as 10–15 min (Anusic et al., 2016a, 2016b). Preliminary evidence suggests DRM measures produce comparable results to ESM (Kahneman et al., 2004).

Only a small number of studies have evaluated the stability of DRM measures (e.g., Anusic et al., 2016a; Krueger & Schkade, 2008), and the longest time interval examined was 4 weeks. Because DRM measures typically focus on a single day in respondents’ lives, important questions remain about the stability of experiences aggregated across such a short window of time. The current study therefore contributes to knowledge about measurement, while also addressing broader theoretical questions about the factors responsible for variation in well-being.

In addition to completing DRM measures of experiential well-being, participants provided ratings of three global well-being variables: life satisfaction, global positive affect, and global negative affect. This allowed us to compare experiential measures of affect to more established global assessments.

## Method

### Participants

We analyzed data from the 2012 to 2014 waves of the approximately nationally representative German Socioeconomic Panel Innovation Sample (GSOEP; Wagner, Frick, & Schupp, 2007; Richter & Schupp, 2015). Participants completed DRM measures once annually in 2012–2014. A total of 2,504 unique participants (52% female; age  $M = 51.78$ ,  $SD = 18.00$ ) provided at least one wave of data (2012:  $n = 2,303$ ; 2013:  $n = 1,920$ ; 2014:  $n = 1,763$ ). On average, participants provided 2.39 waves of data ( $SD = .85$ )—with 1,898 participants (76%) providing at least two waves of data. Attrition analyses revealed that people provided fewer waves of data if, collapsing across waves, they reported greater global happiness ( $r = -.09$ , 95% confidence interval [CI]  $[-0.12, -0.05]$ ), or daily stress ( $r = -.08$ , 95% CI  $[-0.12, -0.04]$ ). No other study variables were associated with total waves provided.<sup>2</sup>

### Measures

**Experiential well-being: DRM positive/negative affect.** At each time, participants systematically reconstructed their prior day. Participants were first asked what time they awoke. Afterward, they were queried, “What did you do next?” Participants selected an activity from a predetermined list (e.g., commuting, socializing) and indicated what time the episode began and ended. This procedure was repeated (i.e., participants were asked, “What did you do next?”) until participants had accounted for their entire day.

Afterward, three of the provided episodes were randomly selected for each participant. For each of these episodes, participants rated the extent to which they felt several emotions *during the episode*: happy, enthusiastic, satisfied, angry, frustrated, sad, worried, and stressed. Each emotion was rated on a scale from 1 (*not at all*) to 7 (*very much*). Having participants rate three randomly selected episodes—rather than every episode (e.g., Kahneman et al., 2004)—dramatically reduces the time required to complete the measure, yet nevertheless appears to produce similar findings (Anusic et al., 2016b).

We formed daily composites for each of the eight emotions by averaging the ratings from the three episodes together. For example, we computed a single “daily happiness” composite for each participant at each wave—which was an average of their reported happiness during the three episodes they had rated.

Because previous research suggests positive and negative affect are independent dimensions (e.g., Watson, Clark, & Tellegen, 1988), we used separate latent variables to capture DRM positive and negative affect at each time point. Daily happiness, enthusiasm, and satisfaction were used as indicators for latent DRM positive affect at each time. Daily anger, frustration, sadness, worry, and stress were used as indicators for latent DRM negative affect at each wave.

Although experiential cognitive well-being (e.g., participants’ momentary appraisals of their circumstances) and experiential affective well-being (e.g., participants’ felt emotions) are theoretically separable (see Figure 1), this distinction is often blurred by the language used in the assessment context. In the GSOEP, momentary satisfaction was explicitly measured as an emotion—rather than an *in vivo* cognitive appraisal of one’s circumstances—and thus we collapsed it together with experiential positive affect ( $\alpha = .85$ ), rather than using it as a single-item measure of experiential–cognitive well-being.

**Global affective well-being.** At each time point, participants rated the extent to which they had generally felt happiness, anger, sadness, and worry over the prior 4 weeks. Each emotion was rated on a scale from 1 (*very seldom*) to 5 (*very often*). We used the happiness question as a single-item indicator of participants’ *global positive affect* at each wave. We used a latent variable to aggregate across the anger, sadness, and worry items to obtain a measure of participants’ *global negative affect* at each wave.

**Global cognitive well-being: Life satisfaction.** Participants’ global *life satisfaction* was assessed each wave using a single item that read, “How satisfied are you with your life, all things considered?” This item was rated on a scale from 0 (*completely dissatisfied*) to 10 (*completely satisfied*). Research suggests single-item measures of life satisfaction have comparable validities to multiitem measures (Cheung & Lucas, 2014).

**Table 1.** Latent Descriptive Statistics and Correlations Across All Time Points.

Latent Variable	M	SD	Latent Correlations With 95% Confidence Intervals				
			1	2	3	4	5
1. DRM positive affect	4.00	0.85	—	[−0.09, 0.03]	[0.27, 0.39]	[0.40, 0.55]	[−0.20, −0.08]
2. DRM negative affect	1.66	0.34	−.03	—	[−0.46, −0.30]	[−0.34, −0.19]	[0.36, 0.54]
3. Life satisfaction	6.43	1.30	<b>.34</b>	<b>−0.38</b>	—	[0.73, 0.89]	[−0.71, −0.57]
4. Global positive affect	3.57	0.52	<b>.49</b>	<b>−0.27</b>	<b>0.81</b>	—	[−0.55, −0.42]
5. Global negative affect	2.74	0.46	<b>−.14</b>	<b>0.45</b>	<b>−0.64</b>	<b>−0.49</b>	—

Note. Latent variables were used to capture the shared variance in each variable across all 3 years. These parameter estimates are the means, standard deviations, and correlations for those *latent variables*. Each latent variable was scaled according to its 2012 indicator. The lower correlation matrix contains the parameter estimates; the upper correlation matrix contains the 95% confidence intervals for the parameter estimates. 95% Confidence intervals for correlations in boldface do not include zero.

## Results

### Latent Correlations Among Study Variables

For all analyses, we used structural equation modeling. Table 1 contains the latent descriptive statistics and correlations for all study variables, collapsing across waves. Specifically, latent variables were created to capture the shared variance in each variable across time (e.g., a single “life satisfaction” latent variable was created, with life satisfaction at 2012, 2013, and 2014 as its indicators). The reported means, standard deviations, and correlations are for these aggregate latent variables.

Theoretically, positive and negative affect are independent (e.g., Watson et al., 1988). Supporting this idea, DRM positive and negative affect were unrelated ( $r = -.03$ , 95% CI [−0.09, 0.03]). In contrast, global positive and negative affect were negatively correlated ( $r = -.49$ , 95% CI [−0.55, −0.42]). This may represent a dissociation in which people believe positive and negative affect are mutually antagonistic, and this potentially inaccurate belief affects their global assessments of their affective experiences (Robinson & Clore, 2002a, 2002b, 2007; cf. Schimmack, 2009).

In terms of convergent validity, DRM positive affect was positively correlated with global life satisfaction ( $r = .34$ , 95% CI [0.27, 0.39]) and global positive affect ( $r = .49$ , 95% CI [0.40, 0.55]), and it was negatively related to global negative affect ( $r = -.14$ , 95% CI [−0.20, −0.08]). The correlations for DRM negative affect mirrored these associations: DRM negative affect was negatively correlated with life satisfaction ( $r = -.38$ , 95% CI [−0.46, −0.30]) and global positive affect ( $r = -.27$ , 95% CI [−0.34, −0.19]), and was positively associated with global negative affect ( $r = .45$ , 95% CI [0.36, 0.54]). These findings align with previous research suggesting DRM measures are valid assessments of well-being that provide information that only partially overlaps with global measures (Anusic et al., 2016a, 2016b).

### Stability in Well-Being Over Time

**Latent test–retest correlations.** We conducted two separate—albeit similar—analyses to examine stability in each well-being variable. First, we computed the latent test–retest correlations for each construct. For constructs with multiple

**Table 2.** Latent Test–Retest Correlations for Well-Being Variables.

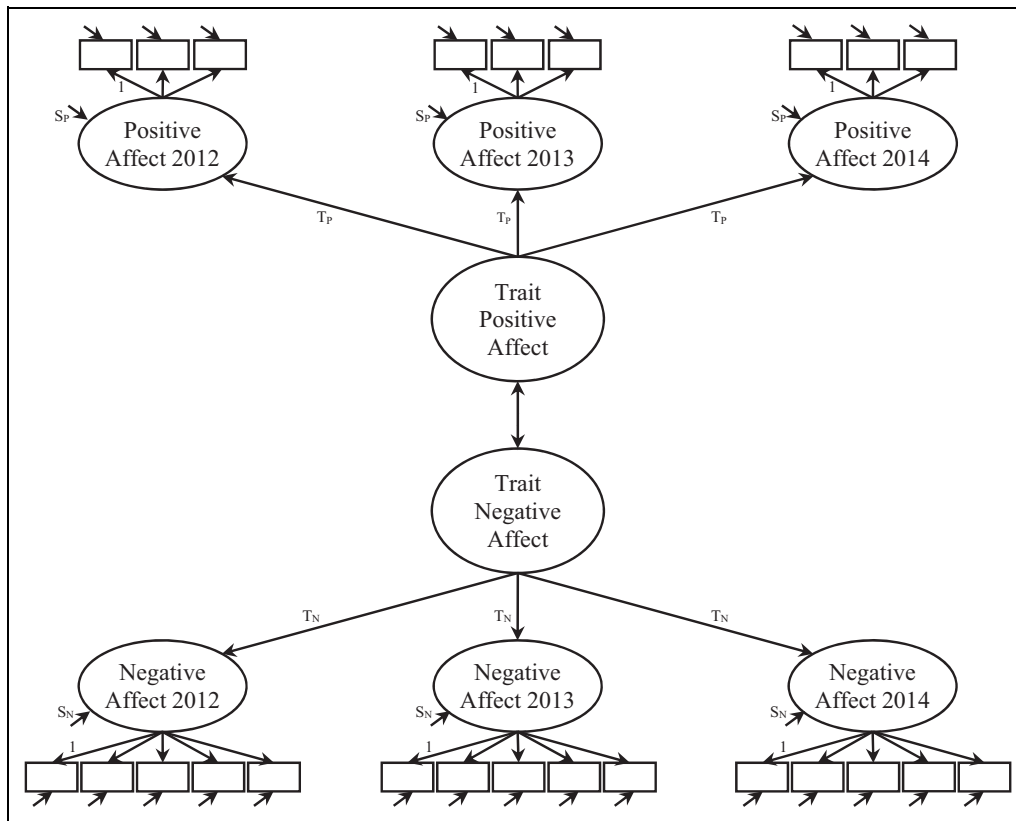
Well-being variable	Times 1–2			Times 2–3			Times 1–3		
	<i>r</i>	LB	UB	<i>r</i>	LB	UB	<i>r</i>	LB	UB
DRM positive affect	<b>.46</b>	.40	.52	<b>.47</b>	.51	.54	<b>.45</b>	.39	.52
DRM negative affect	<b>.30</b>	.25	.36	<b>.39</b>	.33	.46	<b>.32</b>	.26	.38
Life satisfaction <sup>a</sup>	<b>.59</b>	.54	.63	<b>.56</b>	.51	.61	<b>.50</b>	.45	.55
Global positive affect <sup>a</sup>	<b>.44</b>	.39	.49	<b>.48</b>	.43	.52	<b>.44</b>	.39	.48
Global negative affect	<b>.66</b>	.55	.78	<b>.73</b>	.60	.87	<b>.60</b>	.49	.71

Note. 95% CIs for coefficients in boldface do not include zero. DRM = day reconstruction method; CI = confidence interval; LB = lower bound; UB = upper bound.

<sup>a</sup>These were single-item measures; thus the parameters are *manifest* test–retest correlations for these variables.

indicators, latent variables capturing the shared variance across the items *at each time point* were specified. For constructs with a single indicator, manifest variables were used instead. These analyses, presented in Table 2, offer a comparable metric to previous studies that computed test–retest correlations (Krueger & Schkade, 2008) and can be used to evaluate whether the 2-year stabilities (Time 1–Time 3) are lower than the 1-year correlations (Times 1–2; Times 2–3). In general, only global life satisfaction had lower 2-year stability ( $r = .50$ , 95% CI [0.45, 0.55]) than 1-year stability (average  $r = .58$ ). The 2-year stabilities did not differ from the 1-year stabilities for DRM positive affect (2-year:  $r = .45$ , 95% CI [0.39, 0.52]; average 1-year:  $r = .47$ ), DRM negative affect (2-year:  $r = .32$ , 95% CI [0.26, 0.38]; average 1-year:  $r = .35$ ), global positive affect (2-year:  $r = .44$ , 95% CI [0.39, 0.48]; average 1-year:  $r = .46$ ), or global negative affect (2-year:  $r = .60$ , 95% CI [0.49, 0.71]; average 1-year:  $r = .70$ ).<sup>3</sup>

**State–trait models.** Second, we estimated state–trait models using an approach modeled after the Kenny and Zautra framework (e.g., 1995; similar models are described by Khoo, West, Wu, & Kwok, 2006) to determine the percent variance in each well-being measure that was attributable to constant, trait-like



**Figure 2.** State–trait model for day reconstruction method positive and negative affect. Although not depicted for simplicity, residuals for individual indicators were free to covary across time (e.g., the residuals for daily happiness at 2012, 2013, and 2014 were all free to covary with one another).

dynamics across the three measurement occasions.<sup>4</sup> These analyses offer a comparable metric to previous studies that have computed percent variance in well-being attributable to trait- and state-level dynamics (e.g., Anusic et al., 2016a).<sup>5</sup>

As depicted in Figure 2, in state–trait models, a latent variable is used to capture variance shared across all time points. This latent variable is interpretable as the portion of variance attributable to a constant, unchanging, trait-like construct over the course of the study (see Anusic, Lucas, & Donnellan, 2012). The residual terms at each time point are therefore interpretable as the influence of transitory state-level factors. In our specific analyses, we standardized the trait and state latent factors. For parsimony—and to allow the trait to be interpreted as the portion of variance that is *constant* across time—we constrained the trait and state loadings to be equivalent across time. One particularly useful feature of this model specification is the squared standardized loadings represent the portion of variance in the measures that is attributable to constant, trait-like factors, as opposed to malleable state-like factors. Finally, in models containing latent variables (i.e., DRM affect, global negative affect), we allowed the residuals for the indicators to correlate across time (see Cole, Ciesla, & Seiger, 2007).

Table 3 contains estimates of the standardized trait- and state-loadings for each variable ( $\lambda$ ), as well as the percent variance explained in each well-being variable by trait-level and state-level dynamics ( $\lambda^2$ ).<sup>6</sup> Overall, global negative affect was

the most stable study variable over time, with the majority—67% (95% CI [0.59, 0.76])—of its variance attributable to constant, trait-like factors. DRM negative affect was the least stable study variable over time, with 34% (95% CI [0.29, 0.38]) of its variance attributable to trait-level dynamics. Stability was comparable for global life satisfaction ( $\lambda^2 = .56$ , 95% CI [0.51, 0.60]), DRM positive affect ( $\lambda^2 = .46$ , 95% CI [0.41, 0.51]), and global positive affect ( $\lambda^2 = .45$ , 95% CI [0.41, 0.49])—with approximately 50% of the variance in each of these variables attributable to trait-like factors.<sup>7</sup>

## Discussion

Previous research suggests people’s global evaluations of their well-being are relatively stable—even across many years (Schimmack & Oishi, 2005). Although an emerging body of research has begun to examine the stability of experiential well-being, these studies have largely been limited to test–retest intervals of 1 month or shorter (Anusic et al., 2016a; Krueger & Schkade, 2008; cf. Carstensen et al., 2011). In the present study, we examined stability in experiential well-being—as measured via the DRM (Kahneman et al., 2004)—over a span of 2 years. Overall, our findings indicated that when even only a single day of experiences is sampled, a surprisingly large portion of the variance in people’s daily moods and emotions is consistent over a 2-year interval. Indeed, between one-third and

**Table 3.** Trait- and State-Level Variance in Well-Being.

Well-being variable	Trait						State					
	$\lambda$	95% CI		$\lambda^2$	95% CI		$\lambda$	95% CI		$\lambda^2$	95% CI	
		LB	UB		LB	UB		LB	UB		LB	UB
DRM positive affect	<b>.68</b>	.64	.72	<b>.46</b>	.41	.51	<b>.73</b>	.71	.76	<b>.54</b>	.50	.58
DRM negative affect	<b>.58</b>	.54	.62	<b>.34</b>	.29	.38	<b>.81</b>	.78	.84	<b>.66</b>	.61	.71
Global life satisfaction	<b>.75</b>	.72	.77	<b>.56</b>	.51	.60	<b>.67</b>	.65	.68	<b>.44</b>	.43	.47
Global positive affect	<b>.67</b>	.64	.70	<b>.45</b>	.41	.49	<b>.74</b>	.72	.76	<b>.55</b>	.53	.58
Global negative affect	<b>.82</b>	.77	.87	<b>.67</b>	.59	.76	<b>.57</b>	.53	.61	<b>.33</b>	.28	.38

Note. Because of how the model is specified,  $\lambda^2$  represents the proportion of variance in each variable that is attributable to trait- or state-level dynamics. 95% CIs for coefficients in boldface do not include zero. DRM = day reconstruction method; CI = confidence interval; LB = lower bound; UB = upper bound.

one half of the variance in the positive and negative emotions people report on any arbitrary day is due to constant forces across a period of 2 years. We review these findings and their implications in greater depth below.

### Stability in Well-Being

The primary goal of our study was to evaluate the extent to which global and experiential measures of well-being are stable over time. Global negative affect was relatively consistent over 2 years, with two-thirds of its variance attributable to constant, trait-level dynamics. This translates into an average test–retest correlation of  $r = .67$  across the 2-year study duration. In contrast, DRM negative affect was the *least* stable well-being variable—with a mere one-third of its variance due to trait-level dynamics. The remaining well-being variables—life satisfaction, global positive affect, and DRM positive affect—were all approximately 50% explained by constant, trait-level dynamics, and 50% by malleable, state-level processes, which translates into test–retest correlations of approximately  $r = .50$ . This figure aligns with meta-analyses which suggest the average test–retest stability of single-item measures of life satisfaction is approximately  $r = .40$ – $.55$  over a period of 2 years (Anusic & Schimmack, 2016; Schimmack & Oishi, 2005).

Thus on a basic level—and notwithstanding negative affect—our findings indicate global and experiential measures of well-being exhibit similar levels of stability over 2 years. That being said, our results, when integrated with prior research, may suggest different implications for *longer term* stability in global and experiential well-being. Specifically, test–retest correlations tend to attenuate with longer intervals between measurement occasions (e.g., Schimmack & Oishi, 2005). This decay in test–retest stability will, however, asymptotically approach the percent variance in the construct that is *truly* explained by constant, trait-like dynamics (Cole, 2012; Fraley & Roberts, 2005; Fraley et al., 2011). Thus, a phenomenon that is wholly driven by malleable factors (even relatively enduring ones) will eventually exhibit zero test–retest stability over a long enough time interval (which notably, may exceed human lifespan). In contrast, a construct that is driven partially by stable trait-like factors will asymptotically approach some

nonzero test–retest stability, even over indefinitely long intervals.

In studies using state–trait models, the latent trait captures the portion of variance that was constant *across the study's duration*. Thus, any autoregressive (relatively stable, albeit ultimately impermanent) variance that has not fully decayed across the study's duration will be captured as “stable-trait” variance (Anusic et al., 2012). The stable-trait will only represent the portion of variance *truly due to unchanging dynamics* if the study duration is long enough for the measure to reach its asymptotic lower bound in test–retest stability (Anusic et al., 2012; Fraley & Roberts, 2005; Fraley et al., 2011). Thus, our study provides only one piece of the puzzle—the extent to which well-being is stable across 2 years. It must, therefore, be interpreted in conjunction with other studies using varying test–retest intervals to fully understand the extent to which well-being is truly driven by stable, versus slow changing, versus completely transitory factors (Fraley & Roberts, 2005; Fraley et al., 2011).

With respect to global well-being, meta-analyses suggest the test–retest correlation over 2 years is approximately  $r = .40$ – $.50$  for single-item measures of life satisfaction. However, the test–retest correlation decays over time—asymptotically approaching  $r = .20$  or  $.30$  within about 10 years (see Figure 1 in Schimmack & Oishi, 2005; also see Lucas & Donnellan, 2012). Our findings with respect to global well-being align nearly perfectly with this existing body of research and suggest global well-being has trait-like properties.

Far fewer studies have examined the test–retest stability in experiential well-being—especially as measured via DRM. Three previous studies suggest the test–retest correlations for DRM positive affect are approximately  $r = .65$  and  $r = .50$  over a period of 2 and 4 weeks, respectively (Anusic et al., 2016a; Krueger & Schkade, 2008). The present study suggests the 1- and 2-year test–retest reliabilities of experiential positive affect are  $r = .47$  and  $r = .45$ , respectively. Although we strongly caution against drawing conclusions on the basis of four data points, this limited pool of data may tentatively suggest stability in DRM positive affect asymptotically approaches approximately  $r = .40$ – $.50$ —and that it does so as quickly as within a few weeks. This may indicate people's day-to-day positive moods/emotions are generally stable from

1 year to the next. In fact, tentatively extrapolating from the existing trends, stability in experienced positive affect may actually exceed stability in global well-being over a long enough period of time.

In contrast, for DRM negative affect, the 2-week (Krueger & Schkade, 2008), 4-week (Anusic et al., 2016a), 1-year, and 2-year stabilities are approximately  $r = .65, .50, .35,$  and  $.32,$  respectively. This pattern of results may suggest DRM negative affect—at least when aggregated across only a single day—is less stable than DRM positive affect or any global measure of well-being over time (Schimmack & Oishi, 2005). Nevertheless, these inferences are extremely tentative, and far more numerous estimates of the test–retest stability for DRM measures—over varying test–retest intervals—are necessary before strong conclusions can be drawn with respect to their asymptotic test–retest stabilities or the rate at which these variables reach those asymptotic lower bounds (Fraley & Roberts, 2005; Fraley et al., 2011).

### *Implications, Limitations, and Future Directions*

One major implication of the present research is people's day-to-day moods and emotions are remarkably stable. Indeed, up to half of the variance in the affects people feel on any single, arbitrary day is due to stable forces across 2 years. This seems to point to a surprising amount of consistency in the moods and emotions people experience, and provides support that relatively short, efficient, and nonintrusive measures of people's affective experiences can adequately tap enduring individual differences in well-being (Kahneman et al., 2004). This is reassuring for the use of DRM in large-scale surveys, which, due to respondent burden, are unlikely to be able to assess more than a single day of experiences.<sup>8</sup>

That said, the fact we aggregated positive and negative affect across only 1 day at each measurement occasion is also a limitation. Research suggests emotions increase in stability when aggregated over longer periods of time (Diener & Larsen, 1984; Epstein, 1979). Moreover, aggregating greater amounts of emotional data may also potentially increase predictive validities, as well (compare the present Table 1, which aggregates across three DRM occasions with the correlations from Anusic et al., 2016b, which are based on data from only one DRM occasion). The relatively little amount of information aggregated at each time point may have limited our ability to detect stability in or correlates of DRM affect. For example, it may be the case that aggregating across a greater number of days at each measurement occasion would have produced correlations between experiential well-being and life satisfaction that rival the correlations between global affect and life satisfaction. Moreover, increasing the number of days aggregated may further increase the stability of DRM measures—perhaps leading future researchers to conclude experiential measures are more stable and predictive of consequential outcomes than are global measures. Thus, future research should explore the consequences of aggregating across greater numbers of DRM measurement occasions in terms of measurement

reliability, stability, and predictive validity. Doing so should enable scholars to determine an optimal ratio of data collection costs and psychometric benefits.

Relatedly, the use of DRM measures can be viewed as both a strength and limitation of this study. DRM is increasingly being employed in large-scale survey work (e.g., the GSOEP), and our study elucidates that DRM measures can tap stable individual differences in well-being. That said, the DRM measures entail retrospective reporting and may therefore be less valid than ESM (cf. Kahneman et al., 2004).

A third limitation of our study is we did not have measures of experiential cognitive well-being available, as daily satisfaction was explicitly measured as an emotion, rather than momentary appraisal of one's situation.

A final limitation of the present study is we did not have sufficient data to model the temporal dynamics between global and experiential well-being. Scholars are divided with respect to whether people's global reports are less valid than experiential data (e.g., Robinson & Clore, 2002a, 2002b), or whether global and experiential measures validly tap separate stages in a single well-being process, such that changes to one require time to propagate and be reflected in the other (Kim-Prieto et al., 2005). There are several testable implications of the latter perspective. As one example, changes to global and experiential well-being should be corresponsive across time (see Roberts, Wood, & Caspi, 2008). However, the structure of our data—specifically the long duration between waves—was suboptimal for testing such processes. Future research should collect intensive longitudinal data to explore the temporal associations between global and experiential well-being.

### **Conclusion**

Across the entire gamut of society—from individuals to governments—people value their well-being. Previous research has consistently demonstrated people's global evaluations of their well-being are perhaps surprisingly stable over time—with up to half of variance therein due to constant, trait-like dynamics (e.g., Lucas & Donnellan, 2007; Schimmack & Oishi, 2005). Our research suggests people's experiential well-being—the positive and negative moods and emotions they experience on a day-to-day basis—is also relatively stable over 2 years.

### **Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### **Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by a grant from the National Institutes of Health National Institute on Aging (#AG040715).

## Notes

1. Estimates of stability in experience-sampled affect across several days range from intraclass correlation coefficient = .38 (Merz & Roesch, 2011) to cross-time  $\alpha = .95$  (across six measurement occasions; Lischetzke, Angelova, & Eid, 2011).
2. Anusic, Lucas, and Donnellan (2016b) described analyses of experiential affect using only the first wave of this data set. Hudson, Lucas, Donnellan, and Kushlev (2016) examined the associations between income and well-being in this data set.
3. These conclusions are based on whether the average 1-year stability fell within the 2-year stability's confidence interval.
4. We also compared autoregressive-trait/state (ARTS) models to stable-trait/state (STS) models (Anusic et al., 2012). The STS model is nested within the ARTS model, such that autoregressive stability is constrained to unity across time. Thus, the fit of the STS and ARTS models can be compared to determine whether the stability appears to be constant versus decay across time. The less constrained ARTS model did not fit better for DRM positive affect, DRM negative affect, or global positive affect, all  $\chi^2(2)s \leq 4.34$ ,  $p \geq .11$ . This indicates stability in these variables was constant across the study. In contrast, the ARTS model *did* fit better for life satisfaction,  $\chi^2(2) = 31.72$ ,  $p < .01$ , and global negative affect,  $\chi^2(2) = 18.12$ ,  $p < .01$ . These latter findings are ambiguous, however, and may indicate either (1) a lack of stable-trait variance in these constructs (i.e., over a long test–retest interval, their stability would reach zero; Fraley et al., 2011; Fraley & Roberts, 2005), or (2) the present study interval was not sufficiently long for the test–retest estimates to stabilize (Anusic et al., 2012). Thus, the safest conclusion is that stabilities in life satisfaction and negative affect did not reach their asymptotic plateaus—whether zero or non-zero—over 2 years.  
Collectively, these analyses suggest the majority of variables under investigation were not influenced by autoregressive factors across the present study's duration. Thus, in the main text, we use more parsimonious and interpretable trait–state models.
5. Notably, a test–retest correlation of  $r$  implies  $r^2$  of the variance in a construct is due to trait-like dynamics (e.g., a test–retest correlation of  $r = .75$  implies 75% of the variance is attributable to a latent trait).
6. All models fit well, Root Mean Square Error of Approximation (RMSEAs)  $\leq .06$ , Comparative Fit Index (CFIs)  $\geq .93$ .
7. We used multiple-groups models to examine whether stability was invariant among young adults (younger than 40), middle-aged adults (40–59), and older adults (60+). Constraining trait and state variance to be equal across age groups worsened the fits of all models,  $\chi^2(4 \text{ or } 8)s \geq 16.23$ ,  $ps \leq .003$ . Parameter estimates indicated that, aligning with the cumulative continuity principle (Roberts et al., 2008), stable-trait variance was higher among older adults for all variables (average 58.65%) than for middle-aged (average 53.35%) or young adults (average 45.46%).
8. Another related implication of this study is that, counterintuitively, experiential measures may not be more sensitive than global measures to the effects of contextual forces. This may indicate that it is not necessarily advisable to use experiential measures, versus

global ones, in studies designed to examine contextual influences on well-being.

## References

- Anusic, I., Lucas, R. E., & Donnellan, M. B. (2012). Dependability of personality, life satisfaction, and affect in short-term longitudinal data: Dependability in short-term studies. *Journal of Personality, 80*, 33–58.
- Anusic, I., Lucas, R. E., & Donnellan, M. B. (2016a). Comparing the reliability and validity of global self-report measures of subject well-being to day reconstruction measures. *Under Review*.
- Anusic, I., Lucas, R. E., & Donnellan, M. B. (2016b). The validity of the day reconstruction method in the German socio-economic panel study. *Social Indicators Research*. Advance online publication. doi:10.1007/s11205-015-1172-6
- Anusic, I., & Schimmack, U. (2016). Stability and change of personality traits, self-esteem, and well-being: Introducing the meta-analytic stability and change model of retest correlations. *Journal of Personality and Social Psychology, 110*, 766–781.
- Busseri, M. A., & Sadava, S. W. (2011). A review of the tripartite structure of subjective well-being: Implications for conceptualization, operationalization, analysis, and synthesis. *Personality and Social Psychology Review, 15*, 290–314.
- Carstensen, L. L., Turan, B., Scheibe, S., Ram, N., Ersner-Hershfield, H., Samanez-Larkin, G. R., . . . Nesselroade, J. R. (2011). Emotional experience improves with age: Evidence based on over 10 years of experience sampling. *Psychology and Aging, 26*, 21–33.
- Cheung, F., & Lucas, R. E. (2014). Assessing the validity of single-item life satisfaction measures: Results from three large samples. *Quality of Life Research, 23*, 2809–2818.
- Cole, D. A. (2012). Latent trait–state models. In R. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 585–600). New York, NY: Guilford Press.
- Cole, D. A., Ciesla, J. A., & Seiger, J. H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods, 12*, 381–398.
- Diener, E. (1984). Subjective well-being. *Psychological Bulletin, 95*, 542–575.
- Diener, E., & Larsen, R. J. (1984). Temporal stability and cross-situational consistency of affective, behavioral, and cognitive responses. *Journal of Personality and Social Psychology, 47*, 871.
- Diener, E., & Oishi, S. (2004). Are Scandinavians happier than Asians? Issues in comparing nations on subjective well-being. *Asian Economic and Political Issues, 10*, 1–25.
- Dyrenforth, P. S., Kashy, D. A., Donnellan, M. B., & Lucas, R. E. (2010). Predicting relationship and life satisfaction from personality in nationally representative samples from three countries: The relative importance of actor, partner, and similarity effects. *Journal of Personality and Social Psychology, 99*, 690–702.
- Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology, 37*, 1097.
- Fraley, R. C., & Roberts, B. W. (2005). Patterns of continuity: A dynamic model for conceptualizing the stability of individual



- differences in psychological constructs across the life course. *Psychological Review*, *112*, 60–74.
- Fraley, R. C., Vicary, A. M., Brumbaugh, C. C., & Roisman, G. I. (2011). Patterns of stability in adult attachment: An empirical test of two models of continuity and change. *Journal of Personality and Social Psychology*, *101*, 974–992.
- Hudson, N. W., Lucas, R. E., Donnellan, M. B., & Kushlev, K. (2016). Income reliably predicts sadness, but not happiness: A replication and extension of Kushlev, Dunn, & Lucas (2015). *Social Psychological and Personality Science*. Advance online publication. doi: 10.1177/1948550616657599
- Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: The day reconstruction method. *Science*, *306*, 1776–1780.
- Kenny, D. A., & Zautra, A. (1995). The trait-state-error model for multiwave data. *Journal of Consulting and Clinical Psychology*, *63*, 52.
- Khoo, S. T., West, S. G., Wu, W., & Kwok, O. M. (2006). Longitudinal methods. In M. Eid & E. Diener (Eds.), *Handbook of psychological measurement: A multimethod perspective*. Washington, DC: American Psychological Association.
- Kim-Prieto, C., Diener, E., Tamir, M., Scollon, C., & Diener, M. (2005). Integrating the diverse definitions of happiness: A time-sequential framework of subjective well-being. *Journal of Happiness Studies*, *6*, 261–300.
- Krueger, A. B., & Schkade, D. A. (2008). The reliability of subjective well-being measures. *Journal of Public Economics*, *92*, 1833–1845.
- Lischetzke, T., Angelova, R., & Eid, M. (2011). Validating an indirect measure of clarity of feelings: Evidence from laboratory and naturalistic settings. *Psychological Assessment*, *23*, 447–455.
- Lucas, R. E. (2007). Long-term disability is associated with lasting changes in subjective well-being: Evidence from two nationally representative longitudinal studies. *Journal of Personality and Social Psychology*, *92*, 717.
- Lucas, R. E., & Diener, E. (2008). Personality and subjective well-being. In O. John, R. Robins, & L. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 795–814). New York, NY: The Guilford Press.
- Lucas, R. E., Diener, E., & Suh, E. (1996). Discriminant validity of well-being measures. *Journal of Personality and Social Psychology*, *71*, 616.
- 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*, 1091–1098.
- Lucas, R. E., & Donnellan, M. B. (2012). Estimating the reliability of single-item life satisfaction measures: Results from four national panel studies. *Social Indicators Research*, *105*, 323–331.
- Lucas, R. E., & Dyrenforth, P. S. (2006). Does the existence of social relationships matter for subjective well-being? In K. D. Vohs & E. J. Finkle (Eds.), *Self and relationships: Connecting intrapersonal and interpersonal processes* (pp. 245–273). New York, NY: Guilford Press.
- Lucas, R. E., & Schimmack, U. (2009). Income and well-being: How big is the gap between the rich and the poor? *Journal of Research in Personality*, *43*, 75–78.
- Merz, E. L., & Roesch, S. C. (2011). Modeling trait and state variation using multilevel factor analysis with PANAS daily diary data. *Journal of Research in Personality*, *45*, 2–9.
- Okun, M. A., Stock, W. A., Haring, M. J., & Witter, R. A. (1984). The social activity/subjective well-being relation. *Research on Aging*, *6*, 45–65.
- Pinquart, M., & Sorensen, S. (2000). Influences of socioeconomic status, social network, and competence on subjective well-being in later life: A meta-analysis. *Psychology and Aging*, *15*, 187–224.
- Richter, D., & Schupp, J. (2015). The SOEP Innovation Sample (SOEP IS). *Schmollers Jahrbuch*, *135*, 389–400.
- Roberts, B. W., Wood, D., & Caspi, A. (2008). The development of personality traits in adulthood. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 375–398). New York, NY: The Guilford Press.
- Robinson, M. D., & Clore, G. L. (2002a). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, *128*, 934–960.
- Robinson, M. D., & Clore, G. L. (2002b). Episodic and semantic knowledge in emotional self-report: Evidence for two judgment processes. *Journal of Personality and Social Psychology*, *83*, 198–215.
- Robinson, M. D., & Clore, G. L. (2007). Traits, states, and encoding speed: Support for a top-down view of neuroticism/state relations. *Journal of Personality*, *75*, 95–120.
- Samuel, H. (2009). Nicholas Sarkozy wants to measure economic success in “happiness.” *The Telegraph*. Retrieved from <http://www.telegraph.co.uk/news/worldnews/europe/france/6189530/Nicolas-Sarkozy-wants-to-measure-economic-success-in-happiness.html>
- Schimmack, U. (2009). Culture, gender, and the bipolarity of momentary affect: A critical re-examination. *Cognition & Emotion*, *23*, 599–604.
- Schimmack, U., & Oishi, S. (2005). The influence of chronically and temporarily accessible information on life satisfaction judgments. *Journal of Personality and Social Psychology*, *89*, 395–406.
- Schimmack, U., Radhakrishnan, P., Oishi, S., Dzokoto, V., & Ahadi, S. (2002). Culture, personality, and subjective well-being: Integrating process models of life satisfaction. *Journal of Personality and Social Psychology*, *82*, 582–593.
- Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, *4*, 1–32.
- Stratton, A. (2010). Happiness index to gauge Britain’s national mood. *The Guardian*. Retrieved from <http://www.theguardian.com/lifeandstyle/2010/nov/14/happiness-index-britain-national-mood>
- Tay, L., Chan, D., & Diener, E. (2014). The metrics of societal happiness. *Social Indicators Research*, *117*, 577–600.
- University of Waterloo. (2011). *CIW releases Canada’s first-ever index to measure national wellbeing*. Retrieved from <http://uwaterloo.ca/applied-health-sciences/news/ciw-releases-wellbeing-index>
- U.S. Department of Health and Human Services. (2014). *Healthy people 2020*. Retrieved from <http://www.healthypeople.gov/2020/default.aspx>
- Wagner, G. G., Frick, J. R., & Schupp, J. (2007). The German Socio-economic Panel Study (SOEP): Scope, evolution, and enhancements. *Schmollers Jahrbuch*, *127*, 139–169.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063–1070.

### Author Biographies

**Nathan W. Hudson** is a postdoctoral researcher in the Department of Psychology at Michigan State University. His research focuses on adult attachment, close relationships, and adult personality development.

**Richard E. Lucas** is an MSU Foundation Professor in the Department of Psychology of Michigan State University. He studies personality and subjective well-being.

**M. Brent Donnellan** is a professor of psychology in the Social/Personality Program Area at Texas A&M University. His research focuses on personality development across the life span as well as methodological issues related to personality assessment and replication. He is currently the senior associate editor for the *Journal of Research in Personality*.

Handling Editor: Gregory Webster