Supplementary Information

Re-examination of the 10% claim

It turns out that, beyond the simple demographic variables such as age and ethnicity that seem to form the basis of the numbers calculated by Diener and colleagues that were cited by Lyubomirsky, Sheldon, and Schkade (2005), both Andrews and Withey (1976) and Campbell, Converse, and Rodgers (1976) did in fact measure a substantial number of other variables that appear to fall into the category of "life circumstances." Prima facie examples of variables that are candidates for this status include questions such as whether the participant is currently in employment, how many years of education he or she completed, and whether he or she has any health problems[1]. Indeed, these variables appear to meet Lyubomirsky et al.'s own definition of life circumstances:

> This category consists of happiness-relevant circumstantial factors, that is, the incidental but relatively stable facts of an individual's life. Happiness-relevant circumstances may include the national, geographical, and cultural region in which a person resides, as well as demographic factors such as age, gender, and ethnicity . . . Circumstantial factors also include the individual's personal history, that is, life events that can affect his or her happiness, such as having experienced a childhood trauma, being involved in an automobile accident, or winning a prestigious award. Finally, circumstantial factors

---

[1] Of course, as mentioned earlier, a participant's scores on these items might result from an interaction between life circumstances and genetic factors, but for the purpose of our re-analysis we assume that the underlying logic of Lyubomirsky et al.'s (2005) variance decomposition is correct.

include life status variables such as marital status, occupational status, job security, income, health, and religious affiliation. (Lyubomirsky et al., 2005, p. 117)

Thus, both Andrews and Withey's (1976) and Campbell et al.'s (1976) data include a number of variables that ought to allow us to estimate the variance in life satisfaction that these authors might have reported as being explained by life circumstances (versus "classification variables" or "demographic factors"), had this been one of the purposes of their respective studies. Using the original data sets and accompanying documentation for the studies reported by Andrews and Withey (1976) and Campbell et al. (1976), we retained 15 predictors for Andrews and Withey's (1976) May 1972 survey, 15 predictors for the same authors' April 1973 survey, and 18 predictors for Campbell et al.'s 1971 survey. Using software that was a direct descendant of the mainframe programs used by these authors more than 40 years ago, we determined that the percentage of variance explained by either Multiple Classification Analysis, as originally used by Andrew and Withey, or ordinary least squares regression, as used by Campbell et al., was at least R^2 = 18.15% (Andrews & Withey / May), 18.13% (Andrews & Withey / April), and 26.47% (Campbell et al.). That is, had those researchers set out to study the amount of variance explained by life circumstances, they would likely have reported numbers that were on average at least twice as large as Lyubomirsky et al.'s (2005) figure of 10%. Full details of our method, as well as links to all of the information to reproduce our analyses, are provided in the following sections, together with a more conservative re-analysis that uses cross-validation to avoid the dangers of overfitting in models in which all the predictors were treated as categorical variables. This re-analysis also shows that, with an appropriate choice of predictors, the amount of variance robustly predicted by life circumstances can be considerably higher than 10%.

Method

In order to estimate the percentage of variance explained by life circumstances (henceforth LC) in the original studies on which Lyubomirsky et al.'s (2005) article was ultimately based, we downloaded the data sets and accompanying documentation for the study reported by Campbell et al. (1976), and the May 1972 and April 1973 studies reported by Andrews and Withey (1976), from the web site of the Inter-university Consortium for Political and Social Research (ICPSR), as follows:

- For Campbell et al.: https://doi.org/10.3886/ICPSR03508.v1
- For Andrews and Withey: https://doi.org/10.3886/ICPSR03636.v1

We opened the downloaded data files with SPSS, then exported them into comma-separated value (CSV) files for our R code to read. This code then either generated data files for the MicrOSIRIS Multiple Classification Analysis (MCA) software, or performed our equivalent regression analyses (both "classical" OLS and cross-validated) directly.

The data sets that we downloaded contained a large number of variables: 176 from Andrews and Withey's (1976) survey conducted in May 1972, 247 from the same authors' survey conducted in April 1973, and 567 from Campbell et al.'s (1976) survey conducted in 1971, making a total of 990 survey items. However, it was immediately clear that many of these could not possibly be described as measuring LC. Therefore, the first author of the present article conducted an initial coding exercise in which he assigned many of these "obviously not LC" items to one of several categories, as follows:

- "Administrative" (206 variables across the three surveys): items used by the investigators to structure the organization of the studies (e.g., "Week of Interview"). Typically, this information would be determined by the interviewer who was conducting the survey in the participant's home, rather than being provided by the participants themselves.

- "Choice" (26 variables): items that appear to reflect a clear volitional choice by the participant, such as "Do you belong to a sports team?". In terms of the "happiness pie" model, such items fall under the 40% of variance in well-being that Lyubomirsky et al. (2005) appeared to claim can be enhanced with the appropriate choice of activities.
- "Feeling" (323 variables): these are items that reflect participants' expressed opinion or sentiment about some aspect of their lives or their circumstances. Examples are "Are most people trustworthy?," "In what ways is life in the US getting better/worse?," and "How do you feel about your job?" It might be argued that the last of these is to some extent a circumstance; someone who has had a privileged upbringing and become, say, a lawyer might look more favorably on their job than someone who was brought up in poverty and left school early to perform manual labor. However, the subject matter of this item seems to be adequately covered by items such as those asking about pay and promotion chances at the respondent's job, which were put to the panel as candidates to be included as LC.
- "Outcome" (49 variables): items that measure the participant's satisfaction (either with their life overall, or with a particular domain) or some other aspect of their well-being (e.g., "How happy are you these days?"). These are not included because they are the dependent variables that the authors of the original surveys were attempting to explain. Notably, each survey contains one measure of overall life satisfaction that we used as our dependent variable; this was given the specific coding "Focal" to make it easy for our analysis code to pick up.

- “Race–Sex” (6 variables): Lyubomirsky et al. (2005) included “gender” and “ethnicity” in their definition of LC. Hence, we included the variables for these two characteristics in each survey in our analyses.
- “Recode” (128 variables): items for which the responses were not given directly by the participant, but were instead derived by recoding, rescaling, or combining other responses.
- “Uninterpretable” (68 variables): items that did not seem to have any obvious psychological interpretation (e.g., “Country where respondent’s father was born”).

This left 184 variables across the three studies that appeared to be LC candidates. Our aim was, for each study, to further reduce the number of predictors, including only those would be considered as representing LC by a majority of neutral observers. Therefore, as our next step, we recruited six independent coders to participate in a Delphi-like panel (Adler & Ziglio, 1996; Stitt-Gohdes & Crews, 2004) in order to classify the remaining 184 variables into LC or “Other.” Among the coders were one undergraduate, two doctoral candidates, two postdoctoral researchers, and one assistant professor, all working in psychology or related disciplines. Although they were not explicitly told of our hypotheses, they were aware that the purpose of the exercise was to examine whether Lyubomirsky et al.’s (2005) claims about the percentage of variance explained by LC were supported by the available data. They were also given copies of the original code books from the Andrews and Withey (1976) and Campbell et al. (1976) studies, which they could consult if they needed clarification of the meaning of any particular item.

Coding by the Delphi panel proceeded in three rounds. In the first round, each coder indicated, for each variable, whether they considered it to correspond to LC or “Other.” Coders could also leave comments, for example to explain their thought processes in borderline cases.

Once all of the results had been received, the first author collated them. The 34 variables on which there was unanimous agreement were set to one side, and the remainder (150 out of 184) were sent out to the coders for a second round, along with the list of votes from the first round and any comments that had been made. The coders reconsidered and in some cases changed their coding of each variable, and returned their results, which were again collated, resulting in the reduction of the number of variables without unanimous agreement from 150 to 73. (Two coders dropped out of the process after the first round, so the scores in the second and third rounds were based on the results of four people, rather than six.) The process was then repeated for a third round of coding, at the end of which 53 variables remained without unanimous agreement on their coding. In these cases, those with only one dissenting vote ($N = 49$) were treated as if the result had been unanimous, while in the remaining four cases, the tie was broken by the first author, who decided that three variables corresponded to LC and one did not. At all times during this process, all of the coders (who included the second author of the present article) were unaware of the others' identities, except that each coder was identified by the same unique letter throughout the process. The second author of the present article became a co-author late in the coding process; however, she received no more information about the ongoing process of data collection than any other panel member.

At the end of three rounds of coding by the panel, only eight of the 184 variables that had been identified by the first author before the process had been eliminated. It was clear that we needed to reduce this number further, partly to avoid overfitting with classical models (MCA and linear regression) due to an excessive number of predictors, partly to avoid attrition of the effective sample size due to missing data on some items (MCA, in particular, only allows listwise deletion in the case of missing values on a single predictor), and partly because we wanted to establish reasonable *minimum* values for the amount of variance explained by the models that the original authors could have chosen to run, had they set out specifically to investigate the effects of life circumstances (LC) on well-being. We therefore applied a further selection to reduce the number of variables, most notably from Campbell et al.'s (1976) data set, in order to reduce model complexity and to ensure some comparability between the three different samples. This reduced the number of LC variables to 15 (out of 32 that had been identified at the end of the Delphi panel exercise as measuring LC) for Andrews and Withey's (1976) May survey, 15 (out of 35) for the same authors' April survey, and 18 predictors (out of 119) for Campbell et al.'s (1976) survey. Within the retained variables, we also pooled response categories that were very uncommon or appeared to be excessively numerous (for example, we bracketed length of marriage into five-year groups, to avoid having 40 or more categories).

We believe that our process for selecting and recoding LC variables was appropriately conservative. Those decisions taken unilaterally by the authors were always in the direction of reducing the number of predictors; the decisions to include (rather than exclude) variables from the original surveys were taken by majority or unanimous votes by the Delphi panel. None of the three variables for which the first author's casting vote resulted in them having LC status at the end of the Delphi process survived our subsequent simplification. In particular, we accepted that our approach might, on occasion, omit a predictor of well-being that could have been included, under the heading of LC, by another team of researchers. Any such omission would, of course, be likely to result in us underestimating the amount of variance explained.

Having identified a subset of variables that we believed unambiguously corresponded to LC, we then computed the variance explained by those variables in the principal outcome variable for life satisfaction in each study. First, we applied Multiple Classification Analysis

(MCA; Lolle, 2008) with the MicrOsiris software package (Van Eck and Van Eck, 2015). MCA was the technique used exclusively by Andrews and Withey (1976) for their analyses; indeed, MicrOsiris is a direct descendant of the OSIRIS mainframe software used by these authors more than 40 years ago. Second, we applied multiple regression analyses that mirror those used by Campbell et al. (1976). These two analytic approaches (which are almost equivalent, because MCA is essentially a technique for automatically performing the necessary dummy coding to allow categorical predictors in multiple regression) allowed us to answer the question "If these authors had tried to estimate the percentage of variance explained by LC, what numbers would they have found?" Third, since the use of MCA or multiple regression with dummy coded predictors (as performed in the original studies) will likely result in inflated estimates because these highly flexible models might overfit the data on which they are estimated (see Yarkoni & Westfall, 2017, for an introduction to the issue of overfitting), we applied the technique of cross-validation to our regression models to arrive at more realistic/reliable estimates of the percentage of variance that can be explained by LC.

All of the information (R code and other documentation) needed to reproduce our analyses is available at https://osf.io/423sr.

Results

The lists of items measuring LC that were retained at the end of the classification exercise are shown in Table S1. The MicrOsiris software indicated that the percentage of variance explained by the retained variables using MCA was R^2 = 18.15% (Andrews & Withey / May), 18.13% (Andrews & Withey / April), and 26.47% (Campbell et al.). Running multiple regression analyses in R, in which all predictors were treated as categorical variables, resulted in virtually the same estimates, highlighting the conceptual similarity between the MCA approach and the more common multiple regression framework. The percentage of variance explained according to these models were R^2 = 18.19% (Andrews & Withey / May), 18.32% (Andrews & Withey / April), and 26.47% (Campbell et al.). However, because treating all predictors as categorical variables results in extremely flexible models, these numbers might reflect overestimations: Hence, we further reduced the flexibility of the regression models by pooling rare categories and treating predictors such as years of education as continuous variables. We then applied repeated K-fold cross-validation (10 repetitions, 10 folds) to estimate how well the regression models performed on data that had not been used to estimate the model (i.e., *out-of-sample performance*). Using this more conservative approach, the *predicted* variance in life satisfaction was R^2 = 1.92% (Andrews & Withey / May), 6.39% (Andrews & Withey / April), and 17.90% (Campbell et al.). It should be noted that, while the *N*s (1,297, 1,433, and 2,147) of the samples that we re-analyzed exceed the sample sizes of many psychological studies, they are comparably small for the research question at hand, which involves a multitude of potentially intercorrelated predictors, as well as an outcome variable that might be substantially affected by measurement error because of the brief nature of the measurement. This issue could account for the large drop in variance explained when moving from in-sample performance to out-of-sample performance, and highlights the need for large-scale investigations to arrive at reliable and credible estimates.

Discussion

Perhaps somewhat ironically, our most conservative estimates of the variance in well-being explained by LC are even below 10% in two of the three samples. However, the variables included by Campbell et al. (1976) clearly outperformed the 10% figure, even when applying a method that is more conservative than the current domain standard in psychology (Yarkoni &

Westfall, 2017). This can probably be attributed to the fact that Campbell et al.'s study included a broader range of LC, including, for example, social factors such as the reported number of close friends. Once those variables that were essentially identical across the three studies are taken into account, our selection of predictors for the Campbell et al. study includes just six—out of around 100—variables that were unique to that study (i.e., not included by Andrews & Withey, 1976, in either of their studies), suggesting that a few well-chosen variables tapping LC can robustly predict a considerable amount of variance.

<u>Change in the number of within-subject articles in *Journal of Happiness Studies*</u>

On December 19, 2018, we visited the web site of the *Journal of Happiness Studies* and used a null search (https://link.springer.com/search/?facet-journal-id=10902) to retrieve all of the articles from the journal, sorted "Newest First." The first article retrieved was Ingenfeld, Wolbring, and Bless (2018). We selected articles in chronological order down the list until we had obtained 50 abstracts from which we could determine whether the article described an empirical study, and if so, whether within-subject methods were used within that article. After checking 52 abstracts, the last of which was Tavor, Gonen, Weber, and Spiegel (2018) on page 3 of the results, we had obtained 50 empirical articles, of which 16 described within-subject methods. We next sorted the articles "Oldest First" and chose a page that contained articles from 2007–2008 (https://link.springer.com/search/page/12?facet-journal-id=10902&sortOrder=oldestFirst), the first of which was Waterman, Schwartz, and Conti (2008). Again, we moved down the list checking the abstracts (81 in total) until we had obtained 50 empirical articles, the last of which was Collins, Sarkisian, and Winner (2009) on page 16. Of those 50 articles, seven described within-subject methods. A list of all of the articles that we examined can be found at https://osf.io/423sr.

References

Adler, M., & Ziglio, E. (Eds.) (1996). *Gazing into the Oracle*: *The Delphi method and its application to social policy and public health*. London, England: Jessica Kingsley.

Andrews, F. M., & Withey, S. B. (1976). *Social indicators of well-being: Americans' perceptions of life quality*. New York, NY: Plenum.

Campbell, A., Converse, P. E., & Rodgers, W. L. (1976). *The quality of American life: Perceptions, evaluations, and satisfactions*. New York, NY: Russell Sage.

Collins, A. L., Sarkisian, N. & Winner, E. (2009). Flow and happiness in later life: An investigation into the role of daily and weekly flow experiences. *Journal of Happiness Studies*, *10*, 703–719. http://dx.doi.org/10.1007/ s10902-008-9116-3

Ingenfeld, J., Wolbring, T. & Bless, H. (2018). Commuting and life satisfaction revisited: Evidence on a non-linear relationship. *Journal of Happiness Studies*. Advance online publication. http://dx.doi.org/10.1007/s10902-018-0064-2

Lolle, H. (2008). Multiple Classification Analysis (MCA): Unfortunately, a nearly forgotten method for doing linear regression with categorical variables. In *Symposium i anvendt statistik 2008* [Symposium in applied statistics 2008] (pp. 103–122). Copenhagen, Denmark: Copenhagen Business School Press.

Lyubomirsky, S., Sheldon, K. M., & Schkade, D. (2005). Pursuing happiness: The architecture of sustainable change. *Review of General Psychology*, *9*, 111–131. http://dx.doi.org/10.1037/1089-2680.9.2.111

Stitt-Gohdes, W. L., & Crews, T. B. (2004). The Delphi technique: A research strategy for career and technical education. *Journal of Career and Technical Education*, *20*, 55–67. http://dx.doi.org/10.21061/jcte.v20i2.636

Tavor, T., Gonen, L. D., Weber, M., & Spiegel, U. (2018).The effects of income levels and income inequalities on happiness. *Journal of Happiness Studies*, *19*, 2115–2137. http://dx.doi.org/10.1007/s10902-017-9911-9

Van Eck, N., & Van Eck, S. (2015). *MicrOsiris statistical analysis and data management software and decision tree for statistics*. Retrieved March 9, 2018 from https://www.microsiris.com/

Waterman, A. S., Schwartz, S. J. & Conti, R. J. (2008). The implications of two conceptions of happiness (hedonic enjoyment and eudaimonia) for the understanding of intrinsic motivation. *Journal of Happiness Studies*, *9*, 41–79. http://dx.doi.org/10.1007/s10902-006-9020-7

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*, 1100–1122. http://dx.doi.org/10.1177/1745691617693393

Table S1. Survey Questions Included in Our Multiple Classification Analysis and Regression Models.

Andrews & Withey (1976), May survey		Andrews & Withey (1976), April survey		Campbell et al. (1976)	
Q#	Text	Q#	Text	Q#	Text
11	Age of respondent	191	Respondent's age bracket	381	Respondent's age at time of interview
147	Sex of Respondent	192	Respondent's sex	46	Respondent's sex
148	Race of Respondent	185	Respondent's ethnic group	423	Respondent's race
142	Education of Respondent	136	Highest year of school/college completed by respondent	149	What is the highest level of education attained by respondent?
138	Total Family Income Bracket	170	Total family income	270	Respondent's total family income last year
16	Number of children under 18 in respondent's family unit	131	Number of people less than 18 years old in household	24	Number of children in dwelling unit
41	Are you working now, unemployed or laid off, retired and not working, or what?	141	Respondent's present employment status	163	Respondent's current employment status
42	What is your main occupation?	159	Respondent's main occupation	516	Respondent's census occupation classification
44	Do you work for someone else, or yourself, or what?	144	Is respondent employed by someone else, or self-employed?	166	Is respondent self-employed or does s/he work for someone else?
139	Are you married, widowed, divorced, separated, or have you never been married?	178	Respondent's marital status	288	Respondent's marital status
140	How long have you been married?			289	How long has respondent been married?
17	Age of youngest child under 18 in respondent's family unit	231	Age of respondent's youngest child		
18	Age of oldest child under 18 in respondent's family unit				
46	Are you the head of the family unit?	160	Is respondent head of family?		
		171	Does family own house/apartment, or rent it?	108	Does respondent own or rent the dwelling unit?
34	Would you say that you are better off or worse off financially than you were a year ago?				
		180	Do you have a telephone here at home?		
		44	Do you (or anyone else here in your family) own or lease a car?		
				104	How adequate is dwelling unit for respondent and family?
				265	Respondent's religious instruction when growing up
				266	Does respondent have problems with health?
				277	How many good friends does respondent have?
				426	Respondent's apparent intelligence
				387	Did respondent live with father & mother until age 16?

Notes: Q#: Question number from the original survey. Questions on the same line are considered to be equivalent.