Does Smile Intensity in Photographs Really Predict Longevity? A Replication and Extension of Abel and Kruger (2010)

Michael Dufner and Martin Brümmer

University of Leipzig

Joanne M. Chung

Tilburg University

Pia M. Drewke

University of Leipzig

Christophe Blaison

Humboldt-Universität zu Berlin

Stefan C. Schmukle

University of Leipzig

Psychological Science, in press

September 7th 2017, preprint. This article may not exactly replicate the final version published in the journal. It is not the copy of record.

Additional online material and analysis scripts can be retrieved from https://osf.io/8y2ga/

Correspondence: Michael Dufner, Department of Psychology, University of Leipzig, Neumarkt 9-19, 04109 Leipzig, Germany; e-mail: michael.dufner@uni-leipzig.de

Abstract

Abel and Kruger (2010) found that smile intensity, coded from photographs of professional baseball players who were active in the year 1952, predicted these players' longevity. In the current investigation, we sought to replicate this result and to extend the initial analyses. We analyzed (a) a sample that was almost identical to the one from the original study using the same database and inclusion criteria (*N*=224), (b) a considerably larger non-overlapping sample consisting of other players from the same cohort (*N*=527), and (c) all players of the database (*N*=13,530 valid cases). Like Abel and Kruger (2010), we relied on categorical smile codings as indicators of positive affectivity, yet we supplemented these codings with subjective ratings of joy intensity and automatic codings of positive affectivity made by computer programs. In neither sample and for none of the three indicators, positive affectivity predicted mortality once birth year was controlled as a covariate.

Keywords: emotion, affect, longevity, life outcomes, replication

Does Smile Intensity in Photographs Really Predict Longevity? A Replication and Extension of Abel and Kruger (2010)

Past research indicates that dispositional positive affectivity has a life-prolonging function (for a review, see Diener & Chan, 2011). This effect has great theoretical and practical importance. It corroborates the notion that positive affectivity should be regarded as not only an outcome variable but also as a predictor of major life outcomes (Lyubomirsky, King, & Diener, 2005). Moreover, it indicates that policy makers should consider interventions that aim to increase positive affectivity. Furthermore, the effect strengthens personality psychology's standing as a discipline because identifying determinants of longevity is a task of key societal interest (Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007).

Past studies on the topic have assessed positive affectivity by self-report (e.g., Blazer & Hybels, 2004; Lyyra, Törmäkangas, Read, Rantanen, & Berg, 2006), informant-report (Friedman et al., 1993), and content analysis of written text (e.g., Danner, Snowdon, & Friesen, 2001; Pressman & Cohen, 2007). In most studies, however, the timing between assessments of positive affectivity and mortality was short to medium (< 30 years; Diener & Chan, 2011), which is not ideal, because low positive affectivity toward the end of life might be a by-product of illness or physical degradation instead of a genuine predictor of mortality. Only a few studies have investigated time lags of more than four decades, and the results have been contradictory (Danner et al., 2001; Friedman et al., 1993).

In an exceptional study, Abel and Kruger (2010) investigated a sample of professional athletes who were in their prime years of physical fitness and predicted mortality more than five decades later. Specifically, they analyzed photographs of and personal information about U.S. American professional baseball players who were active in the year 1952 and investigated how long they lived up to the year 2009. The authors coded smiling behavior as a proxy for positive affectivity. They did this by classifying smile intensity into three categories (no smile, partial smile, and full smile) and investigated its effects on longevity using a Cox proportional hazards regression model. Baseball players who showed a full (or Duchenne) smile (Ekman, Davidson, & Friesen, 1990) in the photograph were half as likely to die in any given year than players who did not smile (hazard ratio of 0.50). The model included college attendance, marital status, birth year, career length, age at debut year, and body mass index (BMI) as covariates.

In the current research, we revisited the association between smile intensity and longevity by replicating Abel and Kruger's (2010) finding. We relied on the same database as these authors and implemented the same procedures and statistical analyses. We separately analyzed a subsample that was nearly identical to the one analyzed by Abel and Kruger (2010), a non-overlapping subsample consisting of players from the same cohort, and the full database. Like Abel and Kruger (2010), we relied on categorical smile codings as indicators of positive affectivity, but we also supplemented these codings with subjective ratings of joy intensity and automatic codings made by computer programs.

Method

The study was preregistered, and the central parts of the code for data analyses (analyses based on human smile codings and analyses based on automatic codings from one of three emotion recognition computer programs) were uploaded to the internet prior to data collection. Preregistration files, the full data set, and the final code for the current analyses can be downloaded from https://osf.io/8y2ga/.

Sample

We retrieved photographs of all baseball players from the website of the American Baseball Register (http://www.baseball-reference.com/bullpen/Baseball_Register) on September 17, 2015. A total of 18,437 players were listed in the database. For 903 of them, no photograph was available, and therefore, we deleted the data for these players. We also removed 26 cases from the data set for whom information on birth year was missing (and who therefore could not be used in our main analyses). At this stage, the sample size was 17,508.

Data Preparation

We retrieved the same variables as in the original study (death year, college attendance, birth year, career length in the professional league, age at debut year, and BMI).[1] To estimate effects on mortality, we computed two variables: survival status and age. Survival status was determined by whether a player had died (= 1) or had not died (= 0) at the time when we retrieved the data, which was indicated by whether there was a death year in the retrieved data. The age variable indicated the age at which a player died (computed by subtracting the birth year from the death year), or if that the player was still alive, how old the he was at the time of the study (computed by subtracting the birth year from the current year of 2015). When we investigated the distribution of this age variable in the sample of players who were all still alive according to the American Baseball Register, we found that some people had unrealistically high values (see Figure S1 in the online supplemental material). In detail, cases in the distribution ranged up to 100 years, there were no cases between 101 and 125 years, but then there were 42 cases with values equal to or larger than 126 years. The careers of all of these players had ended before 1915, and we deemed it most likely that entries for the variable "death year" were missing for these players from early cohorts. Accordingly, we removed these cases from the data set. Thus, the resulting sample size was 17,466.

Subsamples

In the original study by Abel and Kruger (2010), human codings were obtained for players who were active in 1952, who started their careers before 1950 and who appeared to be looking directly into the camera. There criteria were met for 230 players and for 196 of them all covariates were available, yielding a sample size of 196. We aimed to achieve a sample that was as similar as possible to the original one. Because the judgment of whether or not a player is looking directly into the camera is somewhat vague, we contacted the original authors to obtain access to the original sample. Unfortunately, however, the first author informed us that the original data are not available anymore (E. Abel, personal communication February 21, 2017). Therefore, we applied of the above described criteria by Abel and Kruger (2010) to the data downloaded from the database and had a research assistant judge whether or not players are looking directly into the camera, yielding to a sample size of 224 players. Co-variates were available for all cases. We will refer to this sample as the 1952 sample.

Moreover, we sought to assemble a second subsample from the Baseball Register sample that did not overlap with Abel and Kruger's sample but that was nevertheless as similar as possible. Again, we only included players who were looking into the camera. We followed Simonsohn's (2015) recommendation and aimed to obtain a sample size that was 2.5 times larger than the one used in the original study. Accordingly, approximately half of our human-coding sample consisted of players who ended their career in 1951 (n = 50, all players), 1950 (n = 56, all players), 1949 (n = 68, all players), 1948 (n = 65, all players), or 1947 (n = 28 randomly chosen players). The other half of the sample consisted of players who debuted in 1953 (n = 67, all players), 1954 (n = 63, all players), 1955 (n = 65, all players), 1956 (n = 43, all players), or 1957 (n = 22 randomly chosen players). Thus, our second subsample consisted of 527 players. We will refer to this sample as the non-overlapping replication sample.

Human Codings

As in the original study, a group of five coders (four female, one male; age range: 20 to 35 years) coded the photographs in terms of smile intensity (0 = no smile, 1 = partial smile, 2 = full smile). The operational definitions for these categories were equivalent to the original publication (partial smile = only contraction of the zygomatic major muscles; full smile = movement contraction of both zygomatic and orbicularis oculi muscle). One of the coders (J. M. C.) is a certified coder of the Facial Action Coding System (FACS; Ekman, Friesen, & Hager, 2002). This person taught the other four coders how to distinguish between the three smile categories. Interrater agreement, was Kappa = .61 for both the cases from the 1952 subsample and for the cases from the non-overlapping replication subsample. (In the original study, inter-rater agreement was Kappa = .63.) When raters disagreed, we selected the category that was chosen most frequently. (It never happened that two categories were chosen with equal frequency.) Of the 751 players, 300 (39.95%) showed no smile, 313 (41.68%) showed a partial smile, and 138 (18.37%) showed a full smile.

In addition to these categorical smile codings, which directly matched the design by Abel and Kruger (2010), another group of five raters (undergraduate students; four female, one male; age range: 19 to 20 years) judged the joy intensity shown in the photographs on the basis of their subjective perception (1 = *does not show any joy* to 5 = *shows a lot of joy*). The same 751 players were analyzed as for the smile codings. Interrater agreement was high in both the 1952 sample (ICC = .94) and in the non-overlapping replication sample (ICC = .94), and the subjective joy intensity score averaged across observers was strongly correlated with the categorical codings of smile intensity (Spearman rank-order correlation: $r = .87$, $p < .001$, based on all 751 cases).

Automatic Codings

We relied on three different emotion-recognition computer programs to assess positive affectivity. For theses analyses, we used the complete sample of 17,467 players. However, the results of each program indicated that a number of photographs were uncodable. The number of successfully coded photographs are presented separately for each program below.

First, we used the program FaceReader (Noldus, 2014) version 4.0.8. The program provides continuous scores for happiness, six other emotions, and neutrality, each of which can vary between 0 and 1 (with higher values indicating greater emotional intensity). The program, which has rather strict criteria for identifying a photograph as codable, provided codings for 2,613 cases (2,197 cases for which data on all variables that were relevant for our main analyses were available). Second, we used the emotion-recognition software Emotion

API (Microsoft Cognitive Services, 2016). The program provides probability scores that sum to 1 across happiness, six other emotions, and neutrality. The program provided codings for 15,506 cases (13,531 cases for which data on all relevant variables were available). Third, we used the Computer Expression Recognition Toolbox Version 5.1 (CERT; Littlewort et al., 2011). This program provides a continuous score that indicates smile detection and probability estimates for joy, six other emotions, and neutrality (totaling 1). The program also provides activity scores for the three facial action units (AUs) involved in a full smile (AU6 = “cheek raiser”; AU7 = “lid tightener”; and AU12 = “lip corner puller”). Using CERT, we were able to code 12,417 cases for the smile detection variable (10,652 cases for which data for all relevant variables were available) and 12,419 cases for the remaining variables (10,654 cases for which data on all relevant variables were available).

In addition to treating the positive affectivity scores from the three programs as continuous variables, we also computed dichotomized positive affectivity scores. In each case, a value of 1 was given if positive affectivity predominated over all other emotions, whereas a value of 0 was given if another emotion (or neutrality) predominated over positive affectivity.

Results

Table S1 of the supporting online material shows descriptive statistics for all study variables in the three samples and Table S2 shows the intercorrelations between these variables in the full sample. As can be seen in Table S2, there were substantial correlations between (a) automatic codings of positive affectivity and human codings of smile intensity (values ranged from $r = .70$ to $r = .86$), (b) automatic codings of positive affectivity and human subjective ratings of joy intensity (values ranged from $r = .75$ to $r = .88$, and (c) automatic codings of positive affectivity from the different emotion-recognition programs (values ranged from $r = .62$ to $r = .75$). Correlation coefficients for the activity of the AUs with human codings and ratings of smile intensity varied across the three units and ranged from *r* = .23 to *r* = .71.

Like Abel and Kruger (2010) we used Cox proportional hazards regression models to address our main research question. Such models test the effects of categorical or continuous predictor variables on an event variable (in our case mortality). A significant *b* value for a given predictor indicates that this predictor is linked to mortality. A hazard ratio smaller than 1 means that mortality is less likely with increasing levels of the predictor and a hazard ratio larger than 1 means that mortality is more likely with increasing levels of the predictor. In each subsample and for each operationalization of positive affectivity, we first ran a model to test the effect of positive affectivity on mortality without covariates and then we ran a second model to test the effect of positive affectivity when controlling for the covariates. The results of these main analyses are summarized in Tables 1 and 2.

First, we examined the 1952 subsample. As shown in Table 1, the results for the first model without covariates revealed that mortality could not be predicted by smile intensity, $\chi^2(2) = 0.59$, $p = .746$. The results for the second model with covariates revealed that birth year ($b = -.055$, $SE = .027$, $p = .044$) was a negative predictor of mortality. Thus, players who were born later had a reduced mortality risk. Yet, smile intensity again did not predict mortality, $\Delta\chi^2(2) = .513$, $p = .774$. Neither partial smilers ($b = .102$, $SE = .161$, $p = .529$) nor full smilers ($b = .114$, $SE = .200$, $p = .570$) were less likely to die than people who did not smile. A visual display of these results can be found in Figure S2. When we included subjective joy intensity ratings instead of categorical smile codings as the predictor, joy intensity neither predicted mortality when it was entered as the sole predictor ($b = .005$, $SE = .061$, $p = .929$) nor when it was entered in combination with the covariates ($b = -.011$, $SE = .062$, $p = .862$; see Table S3 for complete model information). Similarly, automatic codings neither predicted mortality when they were entered as sole predictors (see Table S4), nor when they were entered in combination with the covariates (see Table 2).

Table 1

Results of a Cox Regression Analysis Predicting Mortality by Human Smile Codings and Covariates

	1952 SAMPLE					NON-OVERLAPPING REPLICATION SAMPLE				
	b	*SE*	*p*	*HR*	95% CI	*b*	*SE*	*P*	*HR*	95% CI
Model I: Without covariates										
Contrast partial smile vs. non-smile	.063	.158	.689	1.07	[.78, 1.45]	.064	.109	.560	1.07	[.86, 1.32]
Contrast full smile vs. non-smile	.149	.196	.448	1.16	[.79, 1.71]	-.240	.146	.099	.79	[.59, 1.05]
Model II: With covariates										
College	-.269	.159	.090	.76	[.56, 1.04]	-.346	.107	.001	.71	[.57, .87]
Birth year	-.055	.027	.044	.95	[.90, 1.00]	-.029	.009	.001	.97	[.95, .99]
Age at debut	-.028	.036	.443	.97	[.91, 1.04]	-.003	.019	.863	1.00	[.96, 1.04]
Career length	-.040	.025	.111	.96	[.91, 1.01]	-.017	.013	.188	.98	[.96, 1.01]
BMI	.042	.047	.376	1.04	[.95, 1.14]	.054	.037	.144	1.06	[.98, 1.14]
Contrast partial smile vs. non-smile	.102	.161	.529	1.11	[.81, 1.52]	.098	.112	.382	1.10	[.89, 1.37]
Contrast full smile vs. non-smile	.114	.200	.570	1.12	[.76, 1.66]	-.197	.148	.185	.82	[.61, 1.10]

Note. 1952 SAMPLE: N = 224; *SE* = standard error; *HR* = hazard ratio; CI = confidence interval; college attendance: 1 = yes, 0 = no; Model I statistics: $\chi^2(2) = .59$, $p = .75$; Model II statistics: $\chi^2(7) = 10.06$, $p = .19$; the incremental effect of smile codings on mortality in Model II was not significant, $\Delta\chi^2(2) = .51$, $p = .77$; NON-OVERLAPPING REPLICATION SAMPLE: N = 527; *SE* = standard error; *HR* = hazard ratio; CI = confidence interval; college attendance: 1 = yes, 0 = no; Model I statistics: $\chi^2(2) = 4.43$, $p = .11$; Model II statistics: $\chi^2(7) = 32.96$, $p < .001$; the incremental effect of smile codings on mortality in Model II was not significant, $\Delta\chi^2(2) = 4.23$, $p = .12$.

Table 2

Results of Separate Cox Regressions Predicting Mortality by Automatic Codings of Facial Displays of Positive Affectivity from Different Computer Programs (Controlling for Covariates)

	1952 SAMPLE						NON-OVERLAPPING REPLICATION SAMPLE						FULL SAMPLE					
	N	*b*	*SE*	*p*	*HR*	95% CI	*N*	*b*	*SE*	*p*	*HR*	95% CI	*N*	*B*	*SE*	*p*	*HR*	95% CI
FaceReader																		
Happiness	82	.187	.344	.587	1.21	[.61,2.37]	160	-.371	.251	.139	.69	[.42,1.13]	2,197	-.170	.094	.070	0.84	[0.70,1.01]
Happiness dichotomized	82	.146	.262	.576	1.16	[.69,1.93]	160	-.297	.211	.160	.74	[.49,1.12]	2,197	-.105	.074	.158	0.90	[0.78,1.04]
Microsoft Emotion API																		
Happiness	223	.128	.165	.436	1.14	[.82,1.57]	522	-.057	.115	.623	.95	[.75,1.18]	13,530	.040	.034	.241	1.04	[0.97,1.11]
Happiness dichotomized	223	.131	.144	.363	1.14	[.86,1.51]	522	-.017	.103	.871	.98	[.80,1.20]	13,530	.052	.030	.077	1.05	[0.99,1.12]
CERT																		
Smile detection	213	-.002	.016	.891	1.00	[.97,1.03]	480	-.004	.013	.737	1.00	[.97,1.02]	10,652	-.003	.004	.490	1.00	[0.99,1.00]
Joy	213	.173	.196	.379	1.19	[.81,1.75]	480	-.258	.158	.102	.77	[.57,1.05]	10,654	.000	.054	.994	1.00	[0.90,1.11]
Joy dichotomized	213	.110	.152	.468	1.12	[.83,1.5]	480	-.160	.123	.192	.85	[.67,1.08]	10,654	-.001	.041	.971	1.00	[0.92,1.08]
AU6 (“cheek raiser”)	213	.030	.131	.820	1.03	[.80,1.33]	480	-.090	.095	.345	.91	[.76,1.10]	10,654	.013	.029	.663	1.01	[0.96,1.07]
AU7 (“lid tightener”)	213	.002	.343	.996	1.00	[.51,1.96]	480	.044	.230	.850	1.05	[.67,1.64]	10,654	.111	.059	.063	1.12	[0.99,1.25]
AU 12 (“lip corner puller”)	213	.060	.060	.323	1.06	[.94,1.19]	480	-.007	.045	.878	.99	[.91,1.09]	10,654	.005	.013	.714	1.00	[0.98,1.03]

Note. *SE* = standard error; *HR* = hazard ratio; CI = confidence interval; college attendance: 1 = yes, 0 = no.

Second, we examined the non-overlapping replication sample. Table 1 shows that, as in the 1952 subsample, smile codings did not predict mortality when they were included as the sole predictors, $\chi^2(2) = 4.427$, $p = .109$). When smile intensity was entered in combination with the covariates, mortality was significantly predicted by birth year ($b = -.029$, $SE = .009$, $p < .001$) and college attendance ($b = -.346$, $SE = .107$, $p < .001$). However, smile intensity once again did not predict mortality, $\Delta\chi^2(2) = 4.232$, $p = .120$. Neither partial smilers ($b = .098$, $SE = .112$, $p = .382$) nor full smilers ($b = -.197$, $SE = .148$, $p = .185$) were less likely to die than people who did not smile (see Figure S3 for a plot of the results). When we used subjective joy intensity ratings as the predictor variable, they neither predicted mortality when they were entered as the sole predictor ($b = -.047$, $SE = .042$, $p = .264$) nor when they were entered in combination with the covariates ($b = -.037$, $SE = .043$, $p = .390$; see Table S3 for complete model information). When we used automatic codings as predictors, happiness scores from the FaceReader program were negative predictors of mortality ($b = -.493$, $SE = .244$, $p = .043$) when they were entered as sole predictors, but no significant results were present for any of the other automatic codings (see Table S4). When automatically coded variables were entered in combinations with the covariates, none of them predicted mortality (see Table 2).

Third, we the analyzed the automatic codings of positive affectivity in the full sample. When automatic codings of positive affectivity were included as the sole predictors, all of them predicted mortality, and the same was true for activity codings of the three AUs (see Table S4). However, in all cases the effect vanished once we controlled for the covariates (see Table 2 for the main results and Tables S5 to S14 for full model information). This means that automatic codings of positive affectivity did not predict mortality beyond the covariates.

We then explored which covariate was responsible for the initial associations between automatic codings of positive affectivity and mortality in the full sample, and identified birth year as the crucial covariate. Birth year was a consistent negative predictor of mortality in the analyses based on the full sample, and at the same time, it was a negative correlate of positive affectivity displays (see Table S2). Once we controlled for birth year as the sole covariate in our models, the effects of positive affectivity and AU activity became nonsignificant (see Tables S5 to S14). Thus, when we considered that it was uncommon for players from earlier cohorts to smile in photographs (see Figure S4) and that these players also had a reduced life expectancy (see Tables S5 to S14), smiling ceased to predict mortality.

Finally, we considered the possibility that smile intensity might be predictive of mortality solely for specific birth cohorts. To do so, we relied on happiness scores from the Microsoft Emotion API because this program had the least missing values and scores correlated most strongly with human codings. We used happiness scores to predict mortality separately for players who were born prior to 1869, who were born in any specific decade between 1870 and 1989 (1870 to 1879, 1880 to 1889, etc.) and for players who were born after 1970 (again, controlling for all covariates). In none of the analyses happiness was a significant predictor of mortality (see Table S15).

Discussion

Why did the results from the 1952 sample differ from the results reported by Abel and Kruger (2010)? Since we do not have access to the data from the original study, it is not possible to answer this question with certainty. Due to the vagueness inherent in the selection criteria, there were most likely slight differences between the original and the replication sample, which might have led to divergent results. Furthermore, even though agreement among the human coders was reasonably high in both the original and the replication study, each assessment contains a degree of noise, or measurement error, which might also explain the divergent results. The current coders were led by a certified FACS coder, inter-rater agreement was acceptable, and the aggregate coding score correlated substantially with both subjective ratings of joy intensity and automatic codings of positive emotionality. We can thus reasonably conclude that the validity of the current codings was high, but that we are unable to judge the validity of the codings from the original study. Importantly, we did not only fail to detect the expected effect for the categorical smile codings, but also for subjective ratings of joy intensity and automatic codings of positive emotionality.

A similar picture emerged from the analyses that were based on the other samples. The non-overlapping replication sample was substantially larger than the sample of the original study (*N* = 527 vs. *N* = 196), yet nevertheless categorical smile codings, subjective ratings of joy intensity, and automatic codings of positive affectivity all failed to predict mortality when covariates were controlled. In the analyses of the full sample, the *N* was up to more than 69 times larger than the one from the original study. Nevertheless, we again did not find evidence for the premise that smiling has a life-prolonging effect.

When replication studies fail to reproduce an effect, critics often claim that "hidden moderators", such as differences in time, culture, or sample composition between the original and the replication study might account for the null effect (Stroebe & Strack, 2014; Van Bavel, Mende-Siedlecki, Brady, & Reinero, 2016). However, it is particularly difficult to blame such "hidden moderators" in the present case because the sample was drawn from the same population as in the original study, and the cohort that was analyzed was virtually identical. Therefore, in our view, the null results cannot be explained by unassessed moderating factors. Instead, the finding reported by Abel and Kruger (2010) appears to be a false positive result.

The current null results do not imply that dispositional positive affectivity is generally unrelated to longevity. As we stated above, studies relying on different methodologies and analyzing shorter time periods have reported positive links between positive affectivity and longevity (Diener & Chan, 2011). However, the current results indicate that the degree to which professional baseball players smile on a photograph taken during their career does not contain any genuine information about whether they are still alive half a century later.

References

Abel, E. L., & Kruger, M. L. (2010). Smile intensity in photographs predicts longevity. *Psychological Science*, *21*, 542-544. doi: 10.1177/0956797610363775

Blazer, D. G., & Hybels, C. F. (2004). What Symptoms of Depression Predict Mortality in Community-Dwelling Elders?. *Journal of the American Geriatrics Society*, *52*, 2052-2056. doi: 10.1111/j.1532-5415.2004.52564.x

Danner, D. D., Snowdon, D. A., & Friesen, W. V. (2001). Positive emotions in early life and longevity: Findings from the nun study. *Journal of Personality and Social Psychology, 80*, 804–813. doi: 10.1037/0022-3514.80.5.804

Diener, E., & Chan, M. Y. (2011). Happy people live longer: Subjective well-being contributes to health and longevity. *Applied Psychology: Health and Well-Being*, *3*, 1-43. doi: 10.1111/j.1758-0854.2010.01045.x

Ekman, P., Davidson, R. J., & Friesen, W. V. (1990). The Duchenne smile: Emotional expression and brain physiology II. *Journal of Personality and Social Psychology, 58*, 342–353. doi: 10.1037/0022-3514.58.2.342

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *The facial action coding system* (2nd ed.). Salt Lake City, UT: Research Nexus eBook.

Friedman, H. S., Tucker, J. S., Tomlinson-Keasey, C., Schwartz, J. E., Wingard, D. L., & Criqui, M. H. (1993). Does childhood personality predict longevity? *Journal of Personality and Social Psychology*, *65*, 176-185. doi: 10.1037/0022-3514.65.1.176

Littlewort, G., Whitehill, J., Tingfan, W., Fasel, I. R., Frank, M. G., Movellan, J. R., & Bartlett, M. S. (2011). The computer expression recognition toolbox (CERT). IEEE International Conference on Automatic Face & Gesture Recognition and Workshops, 298–305.

Lyrra, T., Törmäkangas, T.M., Read, S., Rantanen, T., & Berg, S. (2006). Satisfaction with present life predicts survival in octogenarians. *Journals of Gerontology. Series B,* *Psychological Sciences and Social Sciences*, *61*, 319–326.doi: 10.1093/geronb/61.6.P319

Lyubomirsky, S., King, L., & Diener, E. (2005). The benefits of frequent positive affect: does happiness lead to success? *Psychological Bulletin, 131*, 803-855. doi: 10.1037/0033-2909.131.6.803

Microsoft Cognitive Services (2015). Retrieved on November 18th and 19th 2015, from https://www.microsoft.com/cognitive-services/en-us/documentation

Noldus. (2014). *FaceReader: Tool for automatic analysis of facial expression: Version 6.0.* Wageningen, the Netherlands: Noldus Information Technology B.V.

Pressman, S. D., & Cohen, S. (2007). Use of social words in autobiographies and longevity. *Psychosomatic Medicine*, *69*, 262-269. doi: 10.1097/PSY.0b013e31803cb919

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, *2*, 313-345. doi: 10.1111/j.1745-6916.2007.00047.x

Simonsohn, U. (2015). Small telescopes detectability and the evaluation of replication results. *Psychological Science*, *26,* 559-569. doi: 10.1177/0956797614567341

Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science*, *9*, 59–71. doi: 10.1177/1745691613514450

Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., & Reiners, D. A. (2016). Contextual sensitivity in scientific reproducibility. *Proceedings of the National Academy of Sciences*, *113*, 6454–6459. doi: 10.1073/pnas.1521897113

Footnote

[1] Marital status was used a covariate in the original but not in the current study because systematic information on players' marital status was lacking when we retrieved the data.