Research Article

Psychological Science
2018, Vol. 29(9) 1451–1462

DOI: 10.1177/0956797618774252
www.psychologicalscience.org/PS
SAGE

"Eavesdropping on Happiness" Revisited: A Pooled, Multisample Replication of the Association Between Life Satisfaction and Observed Daily Conversation Quantity and Quality

Anne Milek[1], Emily A. Butler[1], Allison M. Tackman[1], Deanna M. Kaplan[1], Charles L. Raison[2,3], David A. Sbarra[1], Simine Vazire[4], and Matthias R. Mehl[1]

[1]Department of Psychology, University of Arizona; [2]School of Human Ecology, University of Wisconsin–Madison; [3]School of Medicine and Public Health, University of Wisconsin–Madison; and [4]Department of Psychology, University of California, Davis

Abstract

In the present study, we aimed to replicate and extend findings by Mehl, Vazire, Holleran, and Clark (2010) that individuals with higher well-being tend to spend less time alone and more time interacting with others (e.g., greater conversation quantity) and engage in less small talk and more substantive conversations (e.g., greater conversation quality). To test the robustness of these effects in a larger and more diverse sample, we used Bayesian integrative data analysis to pool data on subjective life satisfaction and observed daily conversations from three heterogeneous adult samples, in addition to the original sample (N = 486). We found moderate associations between life satisfaction and amount of alone time, conversation time, and substantive conversations, but no reliable association with small talk. Personality did not substantially moderate these associations. The failure to replicate the original small-talk effect is theoretically and practically important, as it has garnered considerable scientific and lay interest.

Keywords

Bayesian statistics, happiness, naturalistic observation, well-being, replication, open data, open materials

Received 8/23/17; Revision accepted 4/2/18

It is now a well established and consensually acknowledged fact that social relationships are key to well-being (Argyle, 2001). What is much less clear, however, and what remains a source of considerable debate is whether it is primarily the quantity or the quality of our social encounters that matters. Several gold-standard experience-sampling studies show a linear relationship between how much time people spend interacting with others and how happy they tend to be (Lucas & Dyrenforth, 2006). However, other researchers have argued that the quality of everyday social encounters may be more important than the frequency with which people engage in social contact (e.g., Deci & Ryan, 2000; Myers, 1999).

In a recent study, Mehl, Vazire, Holleran, and Clark (2010) used the Electronically Activated Recorder (EAR; Mehl, 2017) to investigate how well-being relates to observational indicators of real-world conversation quantity and quality. Participants wore the EAR for 4 days while the device intermittently and unobtrusively recorded snippets of ambient sounds as they went about their days. Two measures of interaction quantity (time alone and time in conversation) and quality

Corresponding Authors:
Anne Milek, University of Arizona, Department of Psychology, 1503 E. University Boulevard, Tucson, AZ 85721
E-mail: anne.milek@uzh.ch

Matthias R. Mehl, University of Arizona, Department of Psychology, 1503 E. University Boulevard, Tucson, AZ 85721
E-mail: mehl@email.arizona.edu

(amount of small talk and amount of substantive conversations) were derived from the sampled ambient sounds and linked to well-being. Further strengthening the evidence for the association with quantity, happier participants spent objectively less time alone and more time talking with others. The analyses also provided evidence in favor of quality: Happier participants engaged in more substantive conversations and had a lower ratio of daily small talk accounting for conversation quantity. Importantly, the effects were robust in the sense that they held for weekday and weekend conversations and across multiple measures of well-being.

However, with 79 participants, the sample size, though sizable for labor-intensive naturalistic observation research, was ultimately modest and by current standards too small to yield effects that can be expected to replicate (Maxwell, Kelley, & Rausch, 2008; Schönbrodt & Perugini, 2013).

Beyond statistical power, the generalizability of the sample was also limited in consisting only of college students. Although undergraduate samples are common and valuable in psychology, it is empirically and theoretically important to test whether effects found in a sample that is relatively homogeneous with respect to important social-context variables extend to other populations (Peterson, 2001). For example, undergraduate students are in an environment (i.e., college) that maximally affords getting to know new people and in a life stage where establishing new friendships is normative, relative to middle adulthood, when family and job demands might render the formation of new relationships more difficult and less important for social integration (Havighurst, 1981). In addition, later life stages often include critical life events (e.g., personal illness, trauma or death of a close friend or loved one), which can affect the habitual equilibrium between conversation quantity and quality.

Hence, the question of whether one's well-being is first and foremost a function of the quantity of one's daily conversations or rather a function of their quality remains unanswered in important ways. Well-powered replication studies using age- and context-heterogeneous samples are needed to establish the robustness and generalizability of effect parameters in single studies (Asendorpf et al., 2013). In addition, researchers are increasingly calling for changes in analytical strategies and recommend estimating posterior model probabilities (i.e., a Bayesian approach) in addition to (or instead of) standard frequentist parameters (Carlsson, Schimmack, Williams, & Bürkner, 2017; Johnson, Payne, Wang, Asher, & Mandal, 2017; Marsman et al., 2017). The Bayesian analytical approach seems particularly well suited for replication studies as it "provides a more flexible modeling framework, allows more appropriate quantification of the uncertainty around effect estimates" (Pitchforth & Mengersen, 2012, p. 118), and permits the integration of prior information.

Finally, Mehl and colleagues (2010) took into consideration that participants' personalities might explain differences in well-being (cf. DeNeve & Cooper, 1998; Lucas & Fujita, 2000) by controlling for the Big Five personality domains in their analyses. However, as research shows, participants' personalities exert organizational forces on how individuals act in, and interact with, their social environments (Mehl, Gosling, & Pennebaker, 2006). For example, researchers have suggested that extraverts might benefit more from social interactions than introverts (e.g., Harris, English, Harms, Gross, & Jackson, 2017). Hence, personality might also moderate the link between well-being and conversation quantity and quality, which was not tested in the original study because of limited statistical power.

Therefore, in the present study, we sought to replicate and extend the main findings reported by Mehl and colleagues (2010) and tested whether associations between life satisfaction, as a key component of subjective well-being, and conversation quantity and quality generalize beyond a student sample. To obtain robust overall effect estimates, we used Bayesian integrative data analytic methods by pooling life satisfaction and EAR-observed daily conversation information from three large and diverse samples of working adults in addition to the original data (N = 486). For daily conversation *quantity*, we expected a negative association between life satisfaction and spending time alone as well as a positive association with talking with others. For conversation *quality*, we expected a negative association between life satisfaction and having small talk as well as a positive association with substantive conversations. Given the known link between personality and life satisfaction, replicating the exploratory analyses conducted by Mehl and colleagues (2010), we examined whether the associations between life satisfaction and the conversation variables hold even after controlling for personality. In addition, we tested whether the associations between life satisfaction and daily conversation patterns were moderated by personality but without proposing specific hypotheses (including all Big Five dimensions for the sake of empirical completeness).

Method

In all studies, (a) the total number of excluded observations and the reasons for making those exclusions have been reported, (b) all independent variables have been reported (see also the Supplemental Material available online), (c) all (available) dependent variables or measures that were analyzed for this article's target research question have been reported, and (d) how sample size was determined has been reported.

Participants and procedures

We report how we determined our sample size, all data exclusions, and all available happiness and conversation measures for each study below. Information about the study samples and measures is summarized in Tables 1 through 4.

Study 1. This sample constituted the original sample reported in Mehl and colleagues (2010). Eighty undergraduate students recruited primarily from introductory psychology courses at the University of Texas at Austin completed a series of questionnaires and wore the EAR for 4 days. One participant failed to provide valid EAR data and was excluded from all analyses.

Studies 2a and 2b. As part of a larger study, breast cancer patients and their cohabiting partners were recruited from the Arizona Cancer Center (University of Arizona, Tucson) during regular visits to an oncologist. Couples were eligible if the female partner had a primary diagnosis of Stage I, II, or III breast cancer, they were living together in a marriage-like relationship, and spoke primarily English in their daily conversations. Fifty-six couples gave their consent to participate in the study. Fifty patients (Study 2a) and 51 caregiving partners (Study 2b) provided usable data for the current analyses. Both partners completed a series of questionnaires and wore an EAR device for 3 consecutive days from Friday afternoon to Monday morning (for more details, see Robbins, López, Weihs, & Mehl, 2014).

Study 3. As part of a larger study, 261 medically healthy adults living in Atlanta, Georgia, were recruited by the Emory University Center for Health and Well-Being to participate in a randomized controlled trial of a meditation intervention (ClinicalTrials identifier NCT01643369). A total of 184 participants provided usable data for the current analyses. They completed a series of questionnaires and wore the EAR for 3 days (Friday night through Monday morning) before and after an 8-week meditation intervention (for more details, see Kaplan et al., 2017).

Study 4. As part of a larger study about coping with divorce, 133 adults recently separated from their marital partners were recruited from the larger Tucson, Arizona, area. In all, 122 participants provided usable data for the current analyses by completing sets of questionnaires at Time 1 (initial time of assessment), at Time 2 (3-month follow-up), and at Time 3 (5-month follow-up). At each time of assessment, they also wore the EAR for 3 consecutive days (typically Friday to Monday; for more details, see Hasselmo et al., 2018).

Measures

Life satisfaction. All participants completed the Satisfaction With Life Scale (SWLS; Diener, Emmons, Larsen, & Griffin, 1985), a five-item instrument measuring global evaluations of life satisfaction (from 1, *strongly disagree*, to 7, *strongly agree*). Internal consistency was high in all studies (αs ranged from .84 to .93).

Table 1. Sample Characteristics

Sample characteristic	Study 1	Study 2a	Study 2b	Study 3	Study 4
Population	Undergraduate students	Cancer patients	Spouses of cancer patients	Healthy working adults in a meditation trial	Recently divorced/ separated adults
N	79	50	51	184[a]	122[a]
Female	53.2%	100%	15.7%	66.1%	71.3%
Age	M = 18.70 (SD = 1.41)	M = 56.36 (SD = 14.06)	M = 59.04 (SD = 14.77)	M = 33.48 (SD = 8.39)	M = 43.80 (SD = 10.50)
Ethnicity	66% White, 20% Asian, 11% Hispanic, 3% other	81% White, 12% Hispanic, 4% African American, 4% other/unknown	82% White, 16% Hispanic, 2% Asian	54% White, 31% African American, 7% Asian, 4% Hispanic, 4% other	63% White, 22% Hispanic, 5% African American, 3% Asian, 7% other
EAR sound files per participant	M = 300 (SD = 104)	M = 172 (SD = 59)	M = 180 (SD = 53)	M = 161 (SD = 54)	M = 394 (SD = 118)
EAR sampling rate	30 s every 12.5 min	50 s every 9 min	50 s every 9 min	50 s every 9 min or 30 s every 12.5 min	30 s every 12.5 min

Note: EAR = Electronically Activated Recorder.
[a]For the Big Five measures, *N* differs because of missing values (Study 3: N = 180; Study 4: N = 120).

Table 2. Descriptive Statistics for Key Study Measures

	Study 1			Study 2a			Study 2b			Study 3			Study 4		
Measure	*M*	*SD*	α	*M*	*SD*	α	*M*	*SD*	α	*M*	*SD*	α	*M*	*SD*	α
Life satisfaction	4.52	1.37	.93	5.68	1.19	.84	5.53	1.23	.91	5.04	1.19	.84	4.01	1.51	.91
Extraversion	4.25	1.23	.89	4.92	1.28	.86	4.45	1.02	.80	4.69	1.07	.85	4.78	1.25	.86
Agreeableness	4.98	1.00	.79	5.79	0.69	.69	5.28	0.80	.71	5.56	0.81	.83	5.71	0.80	.77
Conscientiousness	4.50	0.95	.76	5.62	0.99	.83	5.65	0.89	.80	5.43	0.87	.82	5.44	0.91	.79
Neuroticism	3.99	1.16	.82	3.31	1.26	.87	3.11	0.93	.76	3.16	1.02	.86	3.64	1.14	.81
Openness to Experience	5.14	1.06	.85	5.08	1.00	.84	5.27	0.87	.80	5.33	0.78	.77	4.94	0.90	.78

Note: Higher values indicate higher levels of the specific trait (range = 1–7).

Personality. All participants completed the Big Five Inventory (BFI; John, Donahue, & Kentle, 1991), a 44-item instrument assessing personality at the global level of the Big Five personality dimensions (from 1, *strongly disagree*, to 7, *strongly agree*). Internal consistencies for Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience were satisfactory in all studies (αs ranged from .69 to .87).

Observational measures. To obtain an estimate of the objective frequency of participants' daily conversations, we asked participants to wear the EAR from the time they woke up in the morning until they went to bed at night. The EAR is a digital audio recorder that unobtrusively samples daily behavior by intermittently recording snippets of ambient sounds. It captures samples between 5% and 10% of participants' waking hours without them knowing exactly when it is recording. The EAR has been successfully used, with good acceptance and adherence, in various samples that are diverse with respect to age, gender, ethnicity, and location (Mehl & Holleran, 2007). Established privacy-protection and data-confidentiality guidelines (Mehl, 2017; Robbins, 2017) were followed in all studies. Specifically, it was ensured that participants had an opportunity to review their recordings and delete sound files they preferred to remain private. All participants received financial compensation at the end of the study.

The EAR sampling rates varied across studies (30 s or 50 s recordings every 9 min or 12.5 min). For each recording, two trained research assistants (except in Study 1) independently coded whether, in any given sound file, the participant was alone or talking with other people (conversation quantity) and whether a captured conversation was small talk or a substantive conversation (conversation quality). We used the conversational-purpose coding system (adapted from the original study, i.e., Mehl et al., 2010). Codings were mutually exclusive but nonexhaustive; that is, for some conversations, neither category applied (e.g., conversations could also be coded as practical conversations, personal/emotional disclosure, or gossip; for more detail, see Table S1 in the Supplemental Material or the Open Science Framework EAR Repository at https://osf.io/74x3c/). Small talk was defined as an uninvolved, banal conversation in which only trivial information

Table 3. Descriptive Statistics for All Observational Measures

	Study 1			Study 2a			Study 2b			Study 3			Study 4		
Measure	*M*	*SD*	ICC[b]	*M*	*SD*	ICC[c]	*M*	*SD*	ICC[c]	*M*	*SD*	ICC[c]	*M*	*SD*	ICC[c]
Spending time alone	.67	.15	.97	.37	.19	.82	.39	.22	.78	.54	.24	.95	.58	.19	.92
Talking with others	.32	.14	.95	.47	.15	.97	.44	.16	.95	.40	.18	.98	.36	.13	.98
Small talk	.06	.06	.76	.11	.08	.81	.10	.07	.76	.08	.06	.67	.05	.04	.36
Small talk (normalized)[a]	.18	.15		.23	.15		.23	.14		.19	.11		.15	.11	
Substantive conversations	.12	.10	.84	.18	.11	.81	.16	.10	.83	.13	.08	.73	.16	.09	.80
Substantive conversations (normalized)[a]	.36	.25		.36	.15		.34	.14		.32	.14		.43	.15	

[a]To account for individual differences in number of conversations, we computed normalized variables as the number of small talk or substantive conversations relative to a person's total number of recorded conversations. [b]Data were single coded; intercoder reliabilities were computed on the basis of a set of training sound files that all coders coded, and intraclass correlation coefficient (ICC)[2, *k*] reliabilities were computed at the sound-file level. [c]Data were double coded; intercoder reliabilities were computed across the two coders on the aggregated average sound-file level (ICC[1, 2]).

Table 4. Pearson Correlations Between All Study Variables

	Study 1 (N = 79)						Study 2a (N = 50)						Study 2b (N = 51)						Study 3 (N = 184)						Study 4 (N = 122)					
Measure	2	3	4	5	6	7	2	3	4	5	6	7	2	3	4	5	6	7	2	3	4	5	6	7	2	3	4	5	6	7
1. Life satisfaction	**−.36**	**.32**	−.03	−.25	**.26**	.20	.14	.17	.16	−.23	.21	.16	.13	−.06	−.12	−.14	−.05	.06	**−.19**	**.16**	.12	.01	**.22**	**.16**	**−.21**	**.33**	.15	.02	**.29**	.11
2. Spending time alone	—	**−.91**	**−.38**	.09	**−.57**	−.17	—	**−.52**	−.05	.17	**−.30**	−.02	—	**−.66**	−.27	.11	**−.40**	−.06	—	**−.83**	**−.66**	**−.23**	**−.54**	.06	—	**−.82**	**−.35**	−.02	**−.64**	−.15
3. Talking with others		—	**.45**	−.04	**.56**	.07		—	**.29**	−.12	**.69**	.20		—	**.35**	−.22	**.75**	**.29**		—	**.65**	.08	**.74**	.06		—	**.41**	−.01	**.76**	.12
4. Small talk			—	**.78**	−.06	**−.30**			—	**.90**	−.14	**−.41**			—	**.79**	−.10	**−.36**			—	**.73**	**.31**	**−.19**			—	**.85**	−.05	**−.47**
5. Small talk (normalized)				—	**−.32**	**−.45**				—	**−.44**	**−.54**				—	**−.50**	**−.56**				—	**−.15**	**−.28**				—	**−.35**	**−.54**
6. Substantive conversations					—	**.77**					—	**.81**					—	**.81**					—	**.64**					—	**.70**
7. Substantive conversations (normalized)						—						—						—						—						—

Note: All correlations in boldface are significant at $p < .05$, without correcting for multiple comparisons.

was exchanged (e.g., "I stepped on something"; "What are you up to?"). Substantive conversations were defined as involved conversations in which meaningful information was exchanged (e.g., "There are a lot of high stress A-type personalities there," "They have already raised 10 million dollars for Haiti"). Because the data were coded by only one coder in Study 1, intercoder reliabilities were computed on the basis of a set of training sound files that all research assistants coded, and intraclass correlation (ICC)[2, *k*] reliabilities were computed at the sound-file level. For all other studies, ICCs were computed across the two coders on the aggregated, average sound-file level (ICC[1, 2]). ICCs were satisfactory for all categories, ranging from .67 to .98 (Table 3), except for small talk in Study 4 (ICC[1, 2] = .36). For each participant, we converted the EAR codes into relative frequencies (i.e., percentage of valid waking recordings in which a category applied). Because Study 3 was an intervention with implications for conversation behavior, we included only the preintervention data; because Study 4 was an observational study, we averaged across the three assessments to obtain maximally reliable estimates of participants' (general) life satisfaction and conversation patterns. To account for individual differences in the amount of daily conversations, we also computed the percentage of conversations that were small talk or substantive (labeled as *normalized*).

Statistical analysis

Following recommendations for the "new statistics" (Cumming, 2014; Kruschke & Liddell, 2018), we opted against a hypothesis-testing approach (e.g., interpreting *p* values or Bayes factors) in favor of an effect-size-estimation approach. The Bayesian estimation framework appeared particularly suited for a replication analysis, as it allows incorporating prior information and appropriately quantifies the uncertainty around effect estimates. Data, R code, and supplementary analyses are available on the Open Science Framework at osf.io/hp2wx.

First, we conducted a series of Bayesian linear regression analyses (five chains, 10,000 iterations, burn-in period of 1,000 steps) separately for each study and for each predictor using the *brms* R package (Buerkner, 2017). Visual examination of the chain trajectories suggested converging results. To facilitate interpretation and to be compatible with the analyses in the original publication, we transformed posterior Bayesian point estimates of the beta weights as well as the corresponding credible intervals into a correlation metric. The outcome measure was normally distributed, and we therefore used a standard Gaussian regression model. We ran the models using the parameter estimates from Study 1 as priors for Study 2 to Study 4. However, the parameter estimates from the Mehl et al. (2010) study may have been biased because of the small sample size. In the absence of an established literature to guide our choice of priors, we reran the models using noninformative flat priors (the *brms* default) for all estimated parameters. The posterior distributions were relatively robust to changes in the specification of the prior distributions. We therefore report only the set of analyses with flat priors (supplementary analyses including Study 1 priors are provided at osf.io/hp2wx).

Second, to derive overall estimates for the associations between life satisfaction and the EAR-derived variables for conversation quantity and quality, we pooled the raw data of all four samples. Pooling raw data across samples rather than aggregating sample-based summary statistics has the advantage of increasing variability and, hence, statistical power (Scheibehenne, Jamil, & Wagenmakers, 2016; see also Curran & Hussong, 2009). However, when choosing such an approach, one needs to consider nonindependence because participants are nested within studies. Failing to account for between-study variability and nonindependence can lead to biased results (e.g., Carlsson et al., 2017). Conceptually, random-effects models are the gold standard for this type of data structure. Practically, however, with only four studies[1] and in the absence of strong prior knowledge, Bayesian random-effects models failed to reliably converge. Note that our analyses on the pooled data, therefore, assume a fixed effect. To nevertheless account for between-sample sources of variability, and to formally test whether associations differed between studies, we included effect codes for study membership and the corresponding interaction terms (i.e., target predictor by study membership). The Watanabe-Akaike information criterion (WAIC) and leave-one-out (LOO) cross-validation (Vehtari, Gelman, & Gabry, 2017) were used to compare model fit of different models applied to the same data on a conventional scale of "deviance" similar to the Akaike information criterion (AIC; lower WAICs and LOOs denote better model fit). Model comparisons revealed that models with and without interaction terms had a very similar fit, indicating little benefit to allowing effects to vary between studies. This was true for all six conversation variables. Consequently, we excluded the interaction terms from the final models and controlled only for study differences in average life satisfaction.

These final models then also served as reference to test (a) whether participants' personalities accounted for the associations between life satisfaction and daily conversation quantity or quality and (b) whether personality[2] moderated these associations. Theoretically, we were most interested in extraversion as a potential moderator (because of its known association with social relationships and life satisfaction), but we included all Big Five dimensions for empirical completeness.

Finally, to complement the Bayesian approach, we also conducted a traditional frequentist random effects meta-analysis (REMA). We separately calculated effect sizes for all studies and used the Exploratory Software for Confidence Intervals (Cumming, 2014) to obtain the overall maximum likelihood correlation estimates and the corresponding confidence intervals.

Results

Daily conversation quantity

As shown Figure 1a, when considered separately, two out of three new studies (Studies 2 to 4) provided evidence that a negative association between life satisfaction and spending time alone was among the 95% most credible values. When pooled across all studies, the overall posterior Bayesian point estimate of the correlation (r) was −.19, 95% confidence interval (CI) = [−.28, −.09], approximately half the size of the original study. Following the analytic procedures of the original study, we also tested whether personality differences accounted for this correlation. When simultaneously controlling for participants' Big Five scores in the models, we found that the effect of spending time alone was essentially unaffected (r = −.18, 95% CI = [−.26, −.09]). The traditional frequentist REMA produced a mean effect-size r of −.18 (95% CI = [−.34, −.02]).

The sample-specific posterior correlation estimates for well-being and talking with others were positive for all but Study 2b (Fig. 1b). When pooled across studies, the data indicated that life satisfaction was moderately positively associated with talking with others (r = .22, 95% CI = [.13, .31]), less strongly than in the original study. Again, controlling for participants' Big Five scores did not substantially alter the mean posterior correlation estimate (r = .18, 95% CI = [.10, .26]), and the frequentist REMA resulted in a similar mean effect-size r of .24 (95% CI = [.14, .34]).

Daily conversation quality

The sample-specific results for the associations of life satisfaction with small talk (Fig. 2) and with substantive conversations (Fig. 3) demonstrated that posterior Bayesian point estimates and credible intervals varied among the new studies to different degrees for the two conversation quality variables. Posterior Bayesian correlation estimates (rs) for the association between small talk and life satisfaction ranged from −.16 (95% CI = [−.44, .12], Study 2a) to .15 (95% CI = [−.03, .32], Study 4). Similarly, for normalized[3] small talk, a correlation of zero was among the 95% most credible correlations for all but the original study. Results for Studies 2 to 4 remained robust even when the posterior distribution from Study 1 was set as a prior for the regression coefficient. Hence, when collapsed across samples, the most likely posterior point estimates (rs) were .05 (95% CI = [−.04, .14]) for small talk and −.08 (95% CI = [−.17, .01])

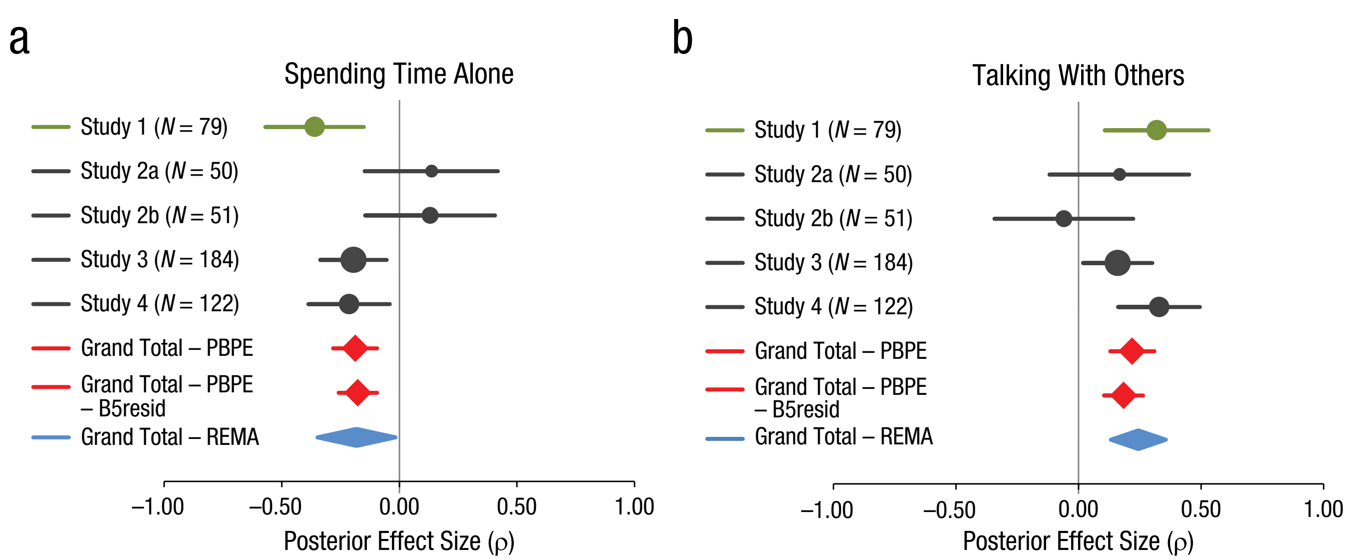

Fig. 1. Posterior Bayesian point estimates for the associations between life satisfaction and (a) spending time alone and (b) spending time talking with others. The posterior Bayesian point estimates (PBPEs, in standardized correlation metric) are indicated as dots or diamonds (size corresponds to sample size), and 95% credible intervals are indicated as horizontal lines. Grand totals include pooled data from Study 1, Study 2a, Study 3, and Study 4 (N = 429). The parallel analysis including data from Study 2b produced similar results. B5resid refers to the effect after we controlled for participants' Big Five personality domain scores. In the random effects meta-analysis (REMA), the effect represents the frequentist maximum likelihood estimate and confidence intervals.

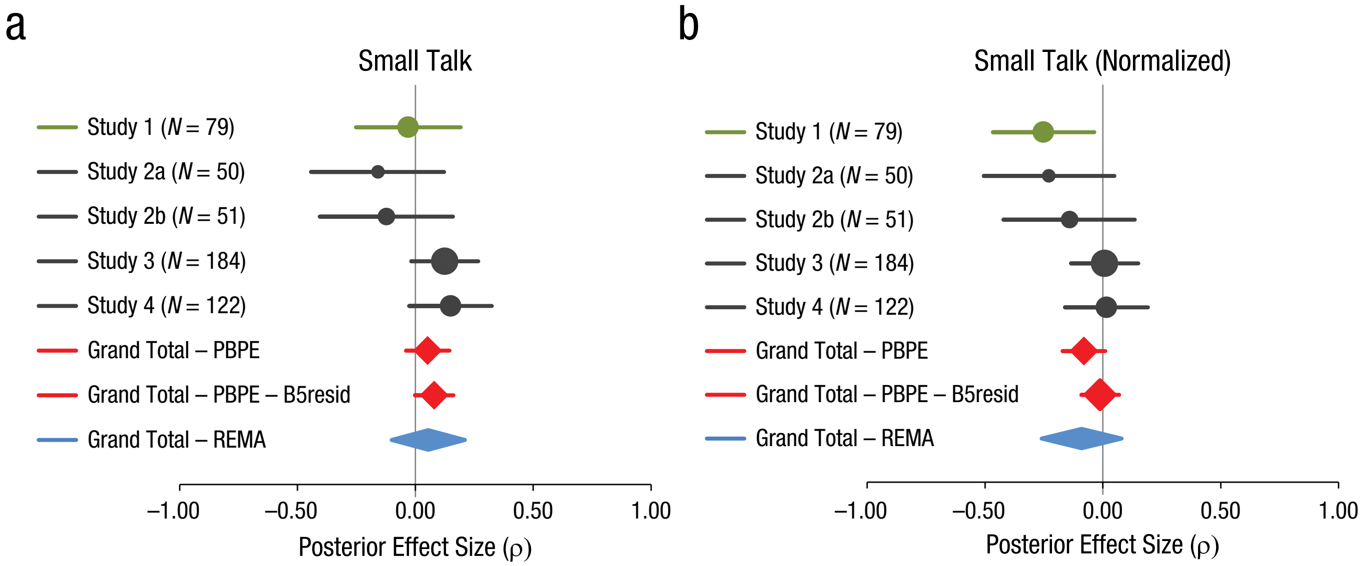

Fig. 2. Posterior Bayesian point estimates for the association between (a) small talk and (b) normalized small talk. The posterior Bayesian point estimates (PBPEs, in standardized correlation metric) are indicated as dots or diamonds (size corresponds to sample size), and 95% credible intervals are indicated as horizontal lines. Grand totals include pooled data from Study 1, Study 2a, Study 3, and Study 4 (N = 429). The parallel analysis including data from Study 2b produced similar results. B5resid refers to the effect after we controlled for participants' Big Five personality domain scores. In the random effects meta-analysis (REMA), the effect represents the frequentist maximum likelihood estimate and confidence intervals.

for normalized small talk. The credible intervals of the two indicators for small talk also contained zero after accounting for the Big Five scores (frequency: r = .08, 95% CI = [−.001, .16]; normalized: r = −.01, 95% CI = [−.09, .07]). The traditional REMA linking life satisfaction to small talk resulted in a mean correlation estimates (rs) of .06 (95% CI = [−.07, .18]) for small talk and −.09 (95% CI = [−.23, .05]) for normalized small talk.

In contrast, posterior correlation estimates for substantive conversations were positive for all but Study 2b. When pooled across studies, the data revealed that life satisfaction was moderately associated with having more substantive conversations (frequency: r = .23, 95% CI = [.14, .32]; normalized: r = .15, 95% CI = [.06, .24]), similar to the original study. The magnitude of the associations was only slightly diminished when we controlled for participants' Big Five scores (frequency: r = .18, 95% CI = [.10, .26]; normalized: r = .10, 95% CI = [.02, .18]). The frequentist REMA linking life satisfaction to substantive conversation yielded effect sizes numerically similar to the Bayesian results (frequency: r = .25, 95% CI = [.15, .34]; normalized: r = .15, 95% CI = [.06, .25]).

Personality as a moderator

Although we refrained from formulating distinct a priori hypotheses about how personality might moderate the observed effects (but, theoretically, were most interested in extraversion as a potential moderator because of its relationship to both participants' social lives and their life satisfaction), we tested whether the slope of the relationship between conversation quantity or quality and life satisfaction differed for people who had different personality traits (e.g., whether introverted people benefitted more or less from small talk than did extraverted people). The additional moderation analyses revealed that individual differences in personality did not substantially moderate the strength of the links between life satisfaction and any of the six conversation variables. Model comparisons using the WAIC and the LOO (Vehtari et al., 2017) revealed that models with interaction terms did not yield better model fit than models without them, providing no convincing evidence for moderation. Either (a) the WAICs or LOOs were substantially smaller in the reference model or (b) the estimated difference (ΔLOOIC) of the expected LOO prediction errors between the moderation model and the reference model included zero in its confidence interval, suggesting no difference in model fit between the models (so the more parsimonious model is to be preferred). In addition, all credible intervals for the posterior point estimates of the interaction terms contained zero (see Tables S2 and S3 in the Supplemental Material). This suggests that, for example, small talk or substantive conversations seemed not to be appreciably more (or less) associated with life satisfaction for extraverted than for introverted participants in our sample.

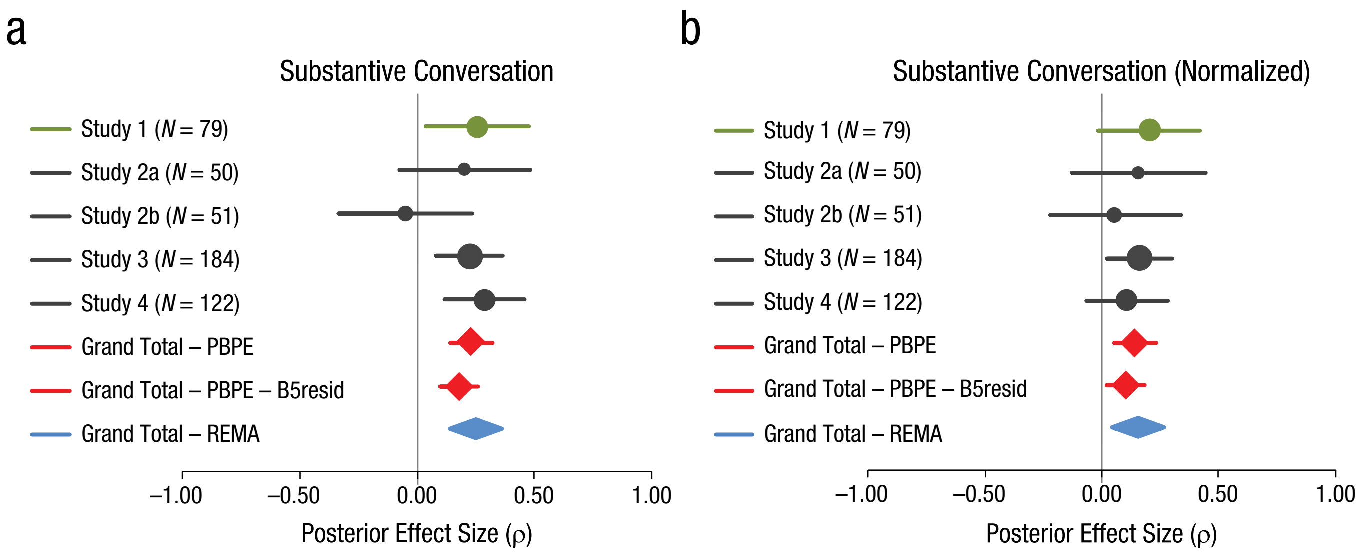

Fig. 3. Posterior Bayesian point estimates for the association between (a) substantive conversation and (b) normalized substantive conversation. The posterior Bayesian point estimates (PBPEs, in standardized correlation metric) are indicated as dots or diamonds (size corresponds to sample size), and 95% credible intervals are indicated as horizontal lines. Grand totals include pooled data from Study 1, Study 2a, Study 3, and Study 4 (N = 429). The parallel analysis including data from Study 2b produced similar results. B5resid refers to the effect after we controlled for participants' Big Five personality domain scores. In the random effects meta-analysis (REMA), the effect represents the frequentist maximum likelihood estimate and confidence intervals.

Discussion

We aimed to replicate and extend previous findings about how conversational properties of social interactions relate to life satisfaction in a heterogeneous adult sample. For daily conversation quantity, we replicated the results reported by Mehl and colleagues (2010) and found evidence for medium-sized associations between life satisfaction and spending time alone (negative) or talking with others (positive), although the negative association for spending time alone was smaller than in the original study.

For daily conversation quality, not all effects were replicated. In accordance with the original findings reported by Mehl and colleagues (2010), we found solid evidence that engaging in substantive conversations was moderately associated with life satisfaction, above and beyond personality characteristics. Our analyses indicated that participants who reported higher life satisfaction than one would expect on the basis of their personality had not only more, but also more substantive, conversations than their less satisfied counterparts with similar personalities.

In contrast, the medium-sized negative association between life satisfaction and the percentage of small-talk conversations—our second indicator of conversation quality—did not replicate: Whether people had more or less small talk in daily life was not reliably associated with life satisfaction. This was true for the overall frequency of small talk as well as the normalized small talk–conversation ratio. The negative correlation suggested by the underpowered original study seems to have been a false-positive finding.

The associations between life satisfaction and conversation quantity and between life satisfaction and substantive conversation were reliably detected in two out of the three new studies. The findings are particularly compelling, as there is no method overlap between the observational conversation codes and the self-report measures of life satisfaction that could explain the links. Mean effect estimates were somewhat smaller than the ones originally reported. This can be expected even in well-powered replication studies and is likely a result of publication bias inflating effect sizes in original studies (Lakens & Etz, 2018). Indeed, given its small sample size and lack of preregistered study and analysis plan, the original study might have been capitalizing on chance, overestimating the true effects. In fact, three out of the four studies—taken individually—were underpowered. This fact underscores the observation that isolated results from small studies should be interpreted with caution and that pooling multiple smaller studies (assuming the absence of a selection bias for inclusion) can be an effective way of increasing power and more reliably estimating population effects (cf. Curran & Hussong, 2009).

The current study further illustrates how the EAR can provide a window into everyday social experiences.

Recently, other research has shown how digital devices (e.g., smartphones, cameras) can successfully be used to collect data to examine psychological, behavioral, and health-related phenomena as they naturally occur in everyday life (e.g., Brown, Blake, & Sherman, 2017; Lathia, Sandstrom, Mascolo, & Rentfrow, 2017). Whereas past research often focused on between-persons comparisons, these new devices are particularly suited to disentangling the within- versus between-persons associations between social activity and well-being. Combining the EAR method with mobile sensing and experience-sampling assessments would enable researchers to compare participants' happiness on days when they have a lot of small talk with days when they predominately engage in substantive conversations. For example, a recent study found evidence for within-persons as well as between-persons links between social interactions and momentary health indicators in daily life (Bernstein, Zawadzki, Juth, Benfield, & Smyth, 2018).

Interestingly and somewhat surprisingly—and in contrast to our expectations specifically with regards to extraversion—we did not find evidence of personality moderating the identified effects. Absence of evidence, however, should not be mistaken for evidence of absence, and future research should follow up on this aspect. Empirically, although our study was well-powered to detect small moderation effects according to conventional power analyses, our effective power may ultimately have been considerably lower considering that it is becoming clear that the field has historically operated on an optimistic consensus of what constitutes small, medium, and large effects (Aguinis, Beaty, Boik, & Pierce, 2005). Hence, it is possible that our study was underpowered to detect small or even medium-sized moderation effects. Theoretically, extraversion might not have emerged as a moderator for the small-talk effect because only the cognitive component of subjective well-being—life satisfaction—was assessed in the current study. Because small talk occurs more often with strangers than with close others, this type of conversation may be an emotionally unpleasant experience for introverts who are reserved but an emotionally pleasant experience for extraverts who are outgoing. Therefore, moderation might be more apparent for the emotional component of well-being that assesses positive and negative affect. Future research should integrate the affective component of well-being, model personality as a latent rather than an observed variable, and examine whether effects of naturalistically observed conversation quantity and quality extend beyond psychological well-being to physical well-being.

The findings reported here should be considered in light of several further limitations. First, even though our explicit goal was to capitalize on the commonalities of four unique samples, results might still be limited with respect to generalizability to the general population. Second, our data are correlational and therefore causally ambiguous: Whether it is the satisfied person who attracts more substantive conversations or whether having substantive conversations makes people more satisfied with their lives is still to be clarified in future (experimental) research. Knowledge about the mechanisms of whether and how substantive conversations relate to better well-being could inform health counselors and inspire innovative interventions. Third, we did not account for the broader social context of the sampled interactions, that is with whom participants were having small talk or substantive conversations (e.g., a stranger vs. a friend). Future research should clarify whether small talk and substantive conversations may relate differentially with well-being in different social or normative contexts. For example, Sandstrom and Dunn (2014) showed that more daily weak-tie interactions (e.g., a small chat with a coffee barista, work colleague, yoga classmate) predicted greater average well-being.

Overall, our findings are consistent with prior intensive longitudinal studies showing that the quantity as well as quality of social interactions matter for well-being (e.g., Carmichael, Reis, & Duberstein, 2015). However, it might be that one of the "active ingredients" of having a conversation is, in fact, that the conversation meets a certain level of meaningfulness. As Baumeister and Leary (1995) argue, mere social contact does not satisfy the basic need to belong. Having close relationships and—as our results suggest—meaningful, substantive conversations might be a key ingredient to a satisfied life.

Action Editor

Brent W. Roberts served as action editor for this article.

Author Contributions

A. Milek and M. R. Mehl developed the study concept. A. Milek and E. A. Butler performed the analyses with input from M. R. Mehl. A. Milek, E. A. Butler, and M. R. Mehl drafted the manuscript. A. M. Tackman, D. M. Kaplan, C. L. Raison, D. A. Sbarra, and S. Vazire provided critical revisions. All the authors approved the final version of the manuscript for submission.

Acknowledgments

The present analyses reproduce the findings reported by Mehl, Vazire, Holleran, and Clark (2010) but otherwise do not overlap with findings reported in other articles published from the included data sets.

Declaration of Conflicting Interests

C. L. Raison has served as a consultant for Alkermes, Novartis, Usona Institute, and Emory Healthcare. C. L. Raison also serves on the steering committee for the North American Center for Continuing Medical Education. The other authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

A. Milek's work on this article was supported by the Swiss National Science Foundation (P2ZHP1_164959). The research was supported by National Institutes of Health Grants 3R01AT004698, 5R01AT004698, R01HD069498, and R03CA137975.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797618774252

Open Practices

All data and materials have been made publicly available via the Open Science Framework and can be accessed at osf.io/hp2wx. The design and analysis plans for the studies were not preregistered. The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797618774252. This article has received badges for Open Data and Open Materials. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

Notes

1. As Study 2 consisted of interdependent couple data, we obtained the overall estimate by first including data only from Study 2a and then replicated all models including data from Study 2b instead. Results did not differ substantially, except for two pooled moderation models; hence, overall estimates from analyses including Study 2a are reported (see the Supplemental Material for results including Study 2b instead of Study 2a).
2. We also considered age and gender as moderators in exploratory analyses (see the Supplemental Material); however, as we did not have specific hypotheses and the associations were not reliably moderated by either of the two, we did not report the effects in the main article.
3. Note that the small talk and normalized small-talk variables are two somewhat different ways of capturing how often a person engages in shallow, banal conversation (similarly, substantive conversations and normalized substantive conversations are two ways to capture how often a person engages in substantive conversation). Whereas the small-talk variable captures the percentage of EAR sound files in which small talk was coded relative to the total number of sound files, the normalized small-talk variable represents the proportion of small talk a participant had as a proportion of all conversations that the EAR captured for this participant.

References

Aguinis, H., Beaty, J. C., Boik, R. J., & Pierce, C. A. (2005). Effect size and power in assessing moderating effects of categorical variables using multiple regression: A 30-year review. *Journal of Applied Psychology*, *90*, 94–107. doi:10.1037/0021-9010.90.1.94

Argyle, M. (2001). *The psychology of happiness* (2nd ed.). New York, NY: Routledge.

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J. A., Fiedler, K., . . .Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality*, *27*, 108–119. doi:10.1002/per.1919

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*, 497–529. doi:10.1037/0033-2909.117.3.497

Bernstein, M. J., Zawadzki, M. J., Juth, V., Benfield, J. A., & Smyth, J. M. (2018). Social interactions in daily life: Within-person associations between momentary social experiences and psychological and physical health indicators. *Journal of Social and Personal Relationships*, *35*, 372–394. doi:10.1177/0265407517691366

Brown, N. A., Blake, A. B., & Sherman, R. A. (2017). A snapshot of the life as lived: Wearable cameras in social and personality psychological science. *Social Psychological and Personality Science*, *8*, 592–600. doi:10.1177/1948550617703170

Buerkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1). doi:10.18637/jss.v080.i01

Carlsson, R., Schimmack, U., Williams, D. R., & Bürkner, P.-C. (2017). Bayes factors from pooled data are no substitute for Bayesian meta-analysis: Commentary on Scheibehenne, Jamil, and Wagenmakers (2016). *Psychological Science*, *28*, 1694–1697. doi:10.1177/0956797616684682

Carmichael, C. L., Reis, H. T., & Duberstein, P. R. (2015). In your 20s it's quantity, in your 30s it's quality: The prognostic value of social activity across 30 years of adulthood. *Psychology and Aging*, *30*, 95–105. doi:10.1037/pag0000014

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*, 7–29. doi:10.1177/0956797613504966

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*, *14*, 81–100. doi:10.1037/a0015914

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, *11*, 227–268. doi:10.1207/S15327965PLI1104_01

DeNeve, K. M., & Cooper, H. (1998). The happy personality: A meta-analysis of 137 personality traits and subjective well-being. *Psychological Bulletin*, *124*, 197–229. doi:10.1037/0033-2909.124.2.197

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment*, *49*, 71–75. doi:10.1207/s15327752jpa4901_13

Harris, K., English, T., Harms, P. D., Gross, J. J., & Jackson, J. J. (2017). Why are extraverts more satisfied? Personality, social experiences, and subjective well-being in college. *European Journal of Personality*, *31*, 170–186. doi:10.1002/per.2101

Hasselmo, K., Mehl, M. R., Tackman, A. M., Carey, A. L., Wertheimer, A. M., Stowe, R. P., & Sbarra, D. A. (2018). Objectively measured social integration is associated with a viral-immune risk profile following marital separation. *Annals of Behavioral Medicine*, *52*, 130–145.

Havighurst, R. J. (1981). *Developmental tasks and education* (3rd ed.). New York, NY: Longman.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory—Versions 4a and 54*. Institute of Personality and Social Research, University of California, Berkeley.

Johnson, V. E., Payne, R. D., Wang, T., Asher, A., & Mandal, S. (2017). On the reproducibility of psychological science. *Journal of the American Statistical Association*, *112*, 1–10. doi:10.1080/01621459.2016.1240079

Kaplan, D., Raison, C., Milek, A., Tackman, A., Pace, T., & Mehl, M. (2017). Mindfulness in daily life: A naturalistic observation study. *PsyArXiv*. doi:10.17605/OSF.IO/R7JCF

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*, 178–206. doi:10.3758/s13423-016-1221-4

Lakens, D., & Etz, A. J. (2018). Too true to be bad: When sets of studies with significant and non-significant findings are probably true. *Social Psychological and Personality Science*, *8*, 875–881. doi:10.1177/1948550617693058

Lathia, N., Sandstrom, G. M., Mascolo, C., & Rentfrow, P. J. (2017). Happier people live more active lives: Using smartphones to link happiness and physical activity. *PLOS ONE*, *12*(1), Article e0160589. doi:10.1371/journal.pone.0160589

Lucas, R. E., & Dyrenforth, P. S. (2006). Does the existence of social relationships matter for subjective well-being? In K. D. Vohs & E. J. Finkel (Eds.), *Self and relationships: Connecting intrapersonal and interpersonal processes* (pp. 254–273). New York, NY: Guilford.

Lucas, R. E., & Fujita, F. (2000). Factors influencing the relation between extraversion and pleasant affect. *Journal of Personality and Social Psychology*, *79*, 1039–1056. doi:10.1037/0022-3514.79.6.1039

Marsman, M., Schönbrodt, F. D., Morey, R. D., Yao, Y., Gelman, A., & Wagenmakers, E.-J. (2017). A Bayesian bird's eye view of "Replications of important results in social psychology." *Royal Society Open Science*, *4*(1), Article 160426. doi:10.1098/rsos.160426

Maxwell, S. E., Kelley, K., & Rausch, J. R. (2008). Sample size planning for statistical power and accuracy in parameter estimation. *Annual Review of Psychology*, *59*, 537–563. doi:10.1146/annurev.psych.59.103006.093735

Mehl, M. R. (2017). The Electronically Activated Recorder (EAR): A method for the naturalistic observation of daily social behavior. *Current Directions in Psychological Science*, *26*, 184–190. doi:10.1177/0963721416680611

Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*, *90*, 862–877. doi:10.1037/0022-3514.90.5.862

Mehl, M. R., & Holleran, S. E. (2007). An empirical analysis of the obtrusiveness of and participants' compliance with the Electronically Activated Recorder (EAR). *European Journal of Psychological Assessment*, *23*, 248–257. doi:10.1027/1015-5759.23.4.248

Mehl, M. R., Vazire, S., Holleran, S. E., & Clark, C. S. (2010). Eavesdropping on happiness: Well-being is related to having less small talk and more substantive conversations. *Psychological Science*, *21*, 539–541. doi:10.1177/0956797610362675

Myers, D. G. (1999). Close relationships and quality of life. In E. Kahnemann, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 374–391). New York, NY: Russell Sage Foundation.

Peterson, R. A. (2001). On the use of college students in social science research: Insights from a second-order meta-analysis. *Journal of Consumer Research*, *28*, 450–461. doi:10.1086/323732

Pitchforth, J. O., & Mengersen, K. L. (2012). Bayesian meta-analysis. In C. L. Alston, K. L. Mengersen, & A. N. Pettitt (Eds.), *Case studies in Bayesian statistical modelling and analysis* (pp. 118–140). Chichester, England: Wiley. doi:10.1002/9781118394472.ch7

Robbins, M. L. (2017). Practical suggestions for legal and ethical concerns with social environment sampling methods. *Social Psychological and Personality Science*, *8*, 573–580.

Robbins, M. L., López, A. M., Weihs, K. L., & Mehl, M. R. (2014). Cancer conversations in context: Naturalistic observation of couples coping with breast cancer. *Journal of Family Psychology*, *28*, 380–390. doi:10.1037/a0036458

Sandstrom, G. M., & Dunn, E. W. (2014). Social interactions and well-being: The surprising power of weak ties. *Personality and Social Psychology Bulletin*, *40*, 910–922. doi:10.1177/0146167214529799

Scheibehenne, B., Jamil, T., & Wagenmakers, E.-J. (2016). Bayesian evidence synthesis can reconcile seemingly inconsistent results: The case of hotel towel reuse. *Psychological Science*, *27*, 1043–1046. doi:10.1177/0956797616644081

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*, 609–612. doi:10.1016/j.jrp.2013.05.009

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*, 1413–1432. doi:10.1007/s11222-016-9696-4