



How to tell a happy person: Accuracy of subjective well-being perception from texts

Olga Stavrova¹ · Lena Haarmann²

Published online: 4 November 2019
© The Author(s) 2019

Abstract

Although perceptions of subjective well-being (SWB) in unacquainted others have been shown to play a major role in impression formation, little is known about how accurate such perceptions are. In two original studies and one pre-registered replication, we explored the accuracy of life satisfaction and happiness judgments from texts and its underlying mechanisms (use of linguistic cues). Participants filled in life satisfaction and happiness measures and completed a brief writing task. Another sample of participants judged the targets' life satisfaction and happiness from the obtained texts. All three studies demonstrated a small to moderate self-other agreement. A linguistic analysis showed that targets with higher (vs. lower) scores on SWB were less likely to use negation words in their texts, which allowed observers to make accurate judgment of their SWB level. Two studies pointed at negative emotion words as valid and positive emotion words as invalid (but often used) cues to happiness, yet these effects did not replicate in Study 3.

Keywords Life satisfaction · Happiness · Accuracy · Positive and negative emotion words

Happiness represents a desired characteristic in romantic partners (Veenhoven 1989) and perceived happiness in others is associated with higher ratings of interpersonal warmth and likeability (Fiske et al. 2007; Rosenberg et al. 1968). But how accurate is our perception of others' happiness? Existing studies have reported a small to moderate association between individuals' self-reports of subjective well-being (SWB: e.g., life satisfaction) and their friends' or relatives' judgment of their SWB (Schneider and Schimmack 2009, 2010). However, the accuracy of the perception of SWB at zero-acquaintance, that is, among strangers, has not received much research attention so far. In fact, given the subjective nature of affective experiences, it has been suggested that affective dispositions, such as SWB, are unlikely to be

judged accurately in a zero-acquaintance context (cf. Watson et al. 2000).

Herein, we sought to test this claim by investigating the accuracy of life satisfaction and happiness judgment at zero acquaintance: specifically, from text samples. Drawing from existing research on the linguistic expression of SWB (Tov et al. 2013) and Brunswik's lens model (Brunswik 1956; Nestler and Back 2013), we proposed that people's SWB level can be accurately detected from texts through a correct utilization of linguistic cues. Across two types of texts—travel reports and creative writing (poetry)—we explored whether independent observers can make accurate ratings of targets' SWB and the utilization of what linguistic cues allows this accuracy.

Judgment accuracy and linguistic cues

Existing research on the accuracy of snap judgments of dispositional characteristics at zero acquaintance has mainly focused on the Big Five personality dimensions. This research has shown that people can relatively accurately identify unacquainted others' Big Five traits based on text samples, such as personal diaries and blogs (Li and Chignell 2010), stream-of-consciousness essays (Holleran and Mehl

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11031-019-09815-4>) contains supplementary material, which is available to authorized users.

✉ Olga Stavrova
O.Stavrova@uvt.nl

¹ Department of Social Psychology, Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands

² University of Cologne, Cologne, Germany

2008), tweets (Qiu et al. 2012a, b), and creative writing (Küfner et al. 2010). Summarizing these findings, a recent meta-analysis (Tskhay and Rule 2014) revealed a small to moderate “accuracy effect”, that is, a small to moderate association between observers’ ratings of targets’ personality and targets’ self-ratings (also referred to as self-other agreement).

In contrast, studies on the accuracy of SWB perception from texts are nearly non-existent (for one exception, see Rodriguez et al. (2010) who showed a high level of accuracy in lay assessment of subclinical depression from individuals’ written self-descriptions), while studies on the accuracy of SWB perception at zero-acquaintance more generally are scarce and have produced mixed findings. For example, people’s judgment of depressive symptoms in unacquainted others based on brief recordings of ambient sounds in their environment (Mehl 2006a) and their Facebook profiles (Fernandez et al. 2012) is no more accurate than chance. In contrast, one study revealed a relatively high accuracy of life satisfaction judgment from brief silent video clips, yet this accuracy effect was restricted to male targets (Yeagley et al. 2007).

Can people accurately judge unacquainted others’ life satisfaction and happiness based on writing samples? According to Brunswik’s lens model (1956), individuals’ psychological characteristics can manifest themselves in observable cues, such as behaviors, choices, facial expressions or language use (Nestler and Back 2013). Brunswik compared such cues to lenses through which observers can see otherwise unobservable psychological traits of individuals. Cues are considered valid, if they are associated with the judged trait. For example, linguistic cues that are indicative of targets’ SWB level are considered valid cues and the association between the observable cue and targets’ score on the underlying trait is referred to as *cue validity*. Observers’ use of cues to make inferences about underlying traits is referred to as *cue utilization*. To make accurate judgments of others’ SWB, observers should be able to differentiate between valid and invalid cues and rely only on those cues that actually reflect targets’ standing on the judged trait. What aspects of individuals’ language use represent valid cues to their happiness?

Decades of research have shown that individual differences in language use can be quite revealing of a range of individuals’ characteristics, including affect (Lee et al. 2007; Lin et al. 2016; Pennebaker and King, 1999; Pennebaker et al. 2003; Yarkoni 2010). For example, dispositionally happy individuals differ from their unhappy counterparts in the way they use positive and negative emotion words. Depressed students are more likely to use negative emotion words in their essays than their non-depressed counterparts (Rude et al. 2004; also see Stirman and Pennebaker 2001). Other studies highlighted the importance of negative (vs.

positive) emotion words as cues to happiness. For example, self-reports of life satisfaction (Liu et al. 2015), depression and anxiety (Settanni and Marengo 2015) correlated with the use of negative, but not positive, emotion words on Facebook. Further research has suggested that, out of impression management concerns, individuals might be more likely to overuse positive emotion words, invalidating them as a happiness indicator (Bazarova et al. 2013; Lin et al. 2014; Qiu et al. 2012a, b).

Even though targets’ use of positive emotion words might not be a valid cue to their happiness, observers might still use it when judging others’ SWB. After all, individuals’ experience of positive emotions is one of the strongest predictors of their own life satisfaction (Kuppens et al. 2008). Therefore, we assumed that targets’ use of both positive and negative emotion words in texts might be seen as valid cues to their happiness by observers.

In brief, in the present research, we examined the role of targets’ use of positive and negative emotion words as the mechanism underlying the accuracy of SWB perception from texts. We assumed that observers would rely on both, positive and negative emotion words, to make judgment of targets’ SWB. However, only negative (but not positive) emotion words will represent valid cues to targets’ SWB.

Overview of the present research

In Study 1, we examined the accuracy of the perception of life satisfaction in travel reports. In Study 2, we complemented these analyses by examining the accuracy of dispositional happiness in creative writing (poetry). Finally, Study 3 provided a preregistered replication of Study 1. Across all studies, we used a linguistic analysis and word count methodology in order to test whether positive and negative emotion words account for the accuracy of SWB judgment from texts. All materials and the data of all studies reported here can be accessed online: https://osf.io/397tq/?view_only=7bfbc78a248745bc83177a69fab41415.

Study 1

Method

Targets

199 individuals completed the study on Amazon Mechanical Turk (MTurk). One gave an implausible answer to one of the questions¹ and one text was not rated due to a programming

¹ This question was a dictator game decision (allocation of \$2 between themselves and another unknown player). It was included for the reasons unrelated to the topic of the present paper. It was com-

error, resulting in a final sample of 197 (mean age = 34.09, 38.6% female) individuals. Participants were asked to complete a writing assignment. Specifically, they were asked to write at least five sentences about their last journey, in particular where they went, how long they stayed, what they liked best and what their usual day looked like. On average, participants wrote 71.93 words ($SD = 36.67$).

To measure life satisfaction, we employed a ladder-measure by Cantril (1965) that asks participants to rate their life overall on a scale ranging from 0 = “the worst possible life overall” to 10 = “the best possible life overall”. Such single-item measures of life satisfaction have good external validity (Cheung and Lucas 2014). Here, we refer to this measure as *self-rated life satisfaction*.

Raters

We recruited another sample of 200 individuals from MTurk to rate the obtained texts. The advertisement of the study was not visible to the MTurk workers who had already participated as “writers”. Six participants gave implausible answers in one of the tasks (s. footnote 1) and were removed, resulting in a final sample of 194 individuals (mean age = 33.61, 47.4% female). Every participant rated five texts. On average, every text was rated by five participants. Raters assessed each target’s life satisfaction using the same measure as the targets did (here referred to as *other-rated life satisfaction*).

Linguistic analysis

To quantify targets’ use of emotion words, we employed LIWC (2007), a widely used and validated text analysis software (Mehl 2006b; Pennebaker et al. 2007; Tausczik and Pennebaker 2010). LIWC uses a word count strategy across different categories, including the frequency of words and word stems denoting affective processes: positive emotion words (e.g. love, nice; 406 words in the dictionary), negative emotion words (e.g. hurt, nasty; 499 words in the dictionary), anger (e.g. hate, annoy; 184 words in the dictionary), anxiety (e.g. fearful, nervous; 91 words in the dictionary) and sadness (e.g. grief, sad; 101 words in the dictionary). These five categories are the only affective word categories included in LIWC. In the present analyses, we focused on these five categories. We also report the correlations with all other linguistic cues in the supplementary materials (Table S1).

Footnote 1 (continued)

pleted by both targets and raters and participants who gave implausible values (e.g., 50\$) were excluded. The materials of the study can be downloaded at: https://osf.io/397tq/?view_only=7bfb78a248745bc83177a69fab41415.

Analytic strategy

Following existing practice, we aggregated observers’ judgment across targets and computed a correlation between mean observer ratings of targets’ SWB and targets’ self-ratings as an indicator of accuracy. In addition, as averaging raters’ judgment of the same target might ignore systematic variation between raters (and can therefore contribute to an increase in Type I error rates), we computed an association between raters’ judgment of each target and each target’s self-judgment in a multilevel regression (Judd et al. 2012; Hox 2002). Multilevel modelling explicitly takes into account the sampling of raters and models the resulting dependencies in the data (e.g., multiple raters judge the same target and multiple targets are judged by the same rater) by partitioning the error into the standard residual error term and random error terms at the level of targets and raters (Judd et al. 2012). As our data have a two-level structure in which each target was rated by multiple raters and each rater rated multiple targets, we chose a cross-classified multilevel model. To examine the degree of accuracy in raters’ judgment of targets’ life satisfaction, we regressed other-rated life satisfaction on self-rated life satisfaction. The intercept was modeled as random at the level of targets and raters. We standardized the variables before the analyses such that obtained coefficients can be interpreted in terms of standard deviations. These analyses were conducted using the lme4 package (Bates et al. 2015) in R.

Results

Accuracy

The correlation between aggregated observer ratings of targets’ life satisfaction and targets’ self-ratings reached $r = .22$ ($p = .002$), showing a moderate accuracy effect. This accuracy effect persisted in a multilevel regression analysis: the effect of self-rated life satisfaction on other-rated life satisfaction reached $b = .14$ ($p < .001$). That is, raters could identify targets’ level of life satisfaction at a better than chance level. A comparison of raters’ to targets’ judgment showed that raters on average overestimated targets’ happiness by .35 on an 11-point scale ($SD = 2.09$; raters: $M = 8.12$, $SD = 1.34$; targets: $M = 7.84$, $SD = 1.92$; $t(196) = 2.34$, $p = .02$).

Linguistic analyses: Cue validity

What linguistic cues distinguished satisfied with life individuals from their less satisfied counterparts? As shown in Table 1, self-rated life satisfaction was negatively associated with the use of anger ($r = -.14$, $p = .050$) and anxiety words ($r = -.15$, $p = .039$). It was not significantly related to other emotion words (all $ps > .10$).

Table 1 Zero-order correlations of linguistic cues with self-rated SWB and other-rated SWB

	Study 1		Study 2		Study 3	
	Cue validity	Cue utilization	Cue validity	Cue utilization	Cue validity	Cue utilization
	Self-rated life satisfaction	Mean other-rated life satisfaction	Self-rated happiness	Mean other-rated happiness	Self-rated life satisfaction	Mean other-rated life satisfaction
Positive emotion words	.09	.15*	.10	.40***	.14*	.002
Negative emotion words	-.08	-.21**	-.26**	-.31**	-.11	-.25***
Anxiety words	-.15*	-.16*	-.09	.01	-.05	-.15*
Anger words	-.14*	-.18*	.01	-.18*	-.07	-.14*
Sadness words	-.01	-.11	-.30**	-.34***	-.06	-.13

Emotion words are percentages from a given category out of the total number of words used. Other-rated life satisfaction and happiness represent average ratings per target

* $p < .05$, ** $p < .01$, *** $p < .001$

Linguistic analyses: Cue utilization

What linguistic cues did observers rely on to make judgment of targets' life satisfaction? Table 1 shows that other-rated life satisfaction was positively associated with targets' use of positive emotion words ($r = .15$, $p = .041$), negatively—with the use of negative emotion words ($r = -.21$, $p = .002$), anger ($r = -.18$, $p = .012$) and anxiety ($r = -.16$, $p = .021$) words.

Mediation analysis

A comparison of cue utilization and cue validity indicators (Table 1) suggests that two out of four linguistic cues that observers relied on were indeed associated with targets' self-rated life satisfaction: anger and anxiety words. To test whether the differential use of anger and anxiety words by satisfied versus less satisfied targets insured a better than chance accuracy in raters' judgment of targets' life satisfaction, we examined whether anger and anxiety words mediated the effect of self-rated life satisfaction on other-rated life satisfaction. Anger and anxiety words were tested as parallel mediators (see Fig. 1). We estimated the effect of the independent variable (self-rated life satisfaction) on the mediators (anger and anxiety words, paths 'a') and the effect of the mediators on the dependent variable (other-rated life satisfaction, paths 'b') in separate multi-level regression equations. To compute the indirect effect, we used the approach recommended by Zhang et al. (2009) for multilevel data. Specifically, we computed the indirect effect by multiplying 'a' and 'b' and used the Monte Carlo estimation method to construct the confidence intervals of the indirect effect (also see Selig and Preacher, 2008). Self-rated life satisfaction was negatively related to the use of anger ($b = -.14$, $p < .001$) and anxiety ($b = -.17$, $p < .001$) words, which in turn predicted observer ratings of targets' life satisfaction ($b = -.10$, $p = .015$, and $b = -.08$, $p = .046$, respectively). Both indirect effects—via anger and anxiety

words—were significant, .014, 95% CI [.002, .028] and .014, 95% CI [.0001, .029], respectively. Hence, less satisfied targets were more likely to use anger- and anxiety-related words in their reports, and independent observers correctly made use of these linguistic cues as indicators of targets' lower life satisfaction.

Discussion

This study showed that individuals can rate strangers' life satisfaction at a better than chance level based on just a few lines of text they wrote. The linguistic analyses showed that to judge targets' life satisfaction, raters extensively relied on targets' use of emotion words, in particular, positive, negative, anger and anxiety words. However, not all of these linguistic cues were helpful in accurately detecting targets' life satisfaction: more satisfied with life targets did not differ from their less satisfied counterparts in their use of positive emotion words and negative emotion words in general. In contrast, words denoting anger and anxiety were associated with both, self- and other-rated life satisfaction and ultimately allowed observers to make accurate judgment of others' life satisfaction.

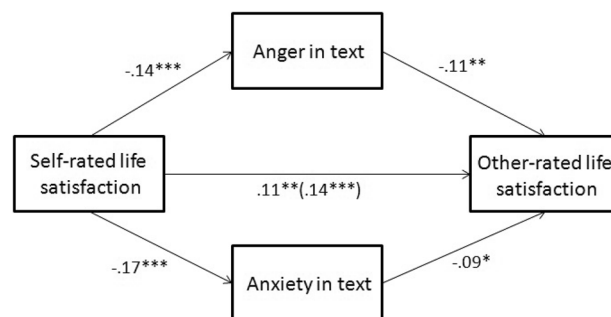


Fig. 1 Mediation analysis, Study 1

Study 2

Study 2 extended these findings to a different writing task (creative writing) and another component of SWB—dispositional happiness that captures individual differences in the tendency to see themselves as a happy or an unhappy person. Like in Study 1, we used a linguistic analysis to examine whether the use of positive and negative emotion words contributes to judgment accuracy.

Method

Targets

151 individuals completed a study on personality and creativity advertised on MTurk. Participants were introduced to a Japanese poetry form—haiku. They read that “a haiku uses just a few words to capture a moment and create a picture in the reader’s mind. It is like a tiny window into a scene. Traditionally, haiku is written in three lines.” We then asked participants to compose a haiku themselves and to feel free to choose any topic they like. We released further requirement (pertaining to the number of syllables etc.), to allow timely study completion and prevent massive drop-out. On average, participants wrote 14.50 words ($SD = 0.71$). Examples of haikus participants composed are: “We throw the peanuts/A squirrel runs by/Fine dining.”; “Leaves on a spring tree/Bright rays of sunlight pass through/To make shadows dance”; “Life passes us by/We are stuck between tomorrow and yesterday/Until we are not any more.”. Fifteen participants didn’t give us the permission to use their haikus in further studies, two individuals composed haikus that consisted of just three words and one haiku was not rated due to a programming error. The data of these individuals were removed from further analyses, resulting in a final sample of 135 (mean age = 35.01, 47.4% female) individuals.

Participants were asked to indicate how accurately the words ‘cheerful’ and ‘sad’ describe them (1 = not at all, 7 = very much). Their responses were combined (“sad” was recoded, $r = .63$, $p < .001$; Cronbach’s $\alpha = .78$) into a measure of *self-rated happiness*.

Raters

Another sample of 678 individuals was recruited on MTurk to rate the obtained haikus. 33 failed a comprehension check question (that required them to select a particular answer instead of answering the question), resulting in a final sample of 645 (mean age = 36.31, 47.9% female) raters. Each participant evaluated three haikus composed by three different targets. On average, every haiku was rated by fourteen

raters. Raters indicated to what extent the words “sad” and “cheerful” describe each target (recoded such that higher values indicate more happiness and combined into a measure of *other-rated happiness*, $r = .70$, $p < .001$; Cronbach’s $\alpha = .82$).

Linguistic analysis

The obtained haikus were analyzed using LIWC following the same procedure as in Study 1.

Results

Accuracy

The correlation between aggregated observer ratings of targets’ happiness and targets’ self-ratings of happiness reached $r = .37$ ($p < .001$). The multilevel analysis revealed that self-rated happiness was positively associated with other-rated happiness ($b = .24$, $p < .001$), replicating the accuracy finding of Study 1. A comparison of raters’ to targets’ judgment showed that raters underestimated targets’ happiness by .62 on a 7-point scale ($SD = 1.38$; raters: $M = 4.41$, $SD = 0.93$; targets: $M = 5.04$, $SD = 1.43$; $t(134) = 5.23$, $p < .001$).

Linguistic analyses: Cue validity

Self-rated happiness was negatively associated with negative emotion ($r = -.26$, $p = .002$) and sadness ($r = -.30$, $p = .001$) words. It was not significantly related to other emotion words (all $ps > .10$; Table 1).

Linguistic analyses: Cue utilization

Other-rated happiness was positively associated with the use of positive emotion words ($r = .40$, $p < .001$), negatively associated with the use of negative emotion words, ($r = -.31$, $p < .001$), sadness ($r = -.34$, $p < .001$) and anger ($r = -.18$, $p = .042$) (Table 1).

Mediation analysis

Like in Study 1, of four cues that observers relied on, only two were actually associated with targets’ self-rated happiness: sadness and negative emotion words. Following the procedure of Study 1, we examined whether the use of sadness and, more generally, negative emotion words mediated the effect of self- on other-rated happiness. However, the category of negative emotion words included sadness (and sadness and negative emotion words were highly correlated: $r = .70$, $p < .001$). Therefore, to be able to determine whether the use of negative emotion words explains variance in happiness beyond and above the use of sadness words, we

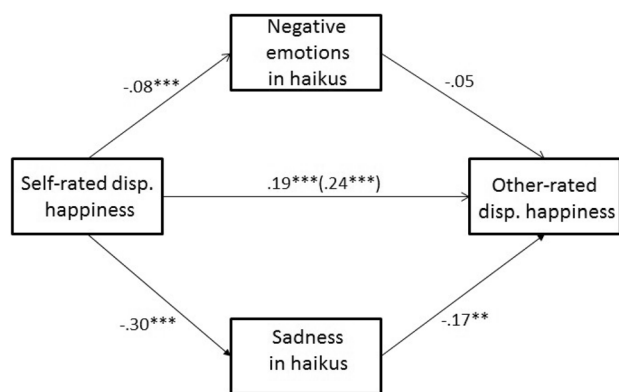


Fig. 2 Mediation analysis via negative emotion (other than sadness) and sadness words, Study 2

excluded sadness words from the negative emotion words count (specifically, we subtracted the sadness words from the overall negative emotion words). We then tested both sadness and negative emotion words as parallel mediators.

Targets' self-rated happiness was negatively related to their use of sadness ($b = -.30, p < .001$) and negative emotion words ($b = -.08, p < .001$) in haikus. Yet, only the use of sadness (but not negative emotion words: $b = -.05, p = .30$) was significantly associated with observer ratings of targets' happiness ($b = -.17, p = .002$) (Fig. 2). The indirect effect via the use of sadness words was significant (.05, 95% CI [.02; .08]), whereas the indirect effect via the use of negative emotion words was not ($-.004, 95\% \text{ CI } [-.01; .004]$).

Additional exploratory analyses

Moderation by topic valence Does the diagnosticity of emotion words in texts depend on the general valence of the text topic? For example, positive emotion words may be particularly revealing in texts written on a negative topic, and negative emotion words—in texts written on a positive topic. To answer this question, we rated each haiku's topic on the dimension of valence (1 = positive, 0 = neutral, -1 = negative). Both authors rated all haikus independently and were blinded to self- and other-rated happiness scores associated with the haikus (and any other information about the haikus). Most haikus were categorized as having a neutral (rater 1: 53.3%; rater 2: 34.1%) or a positive (rater 1: 26.7%; rater 2: 45.2%) topic; and only about one-fifth of the haikus were categorized as having a negative topic (rater 1: 20.0%; rater 2: 20.7%). The raters reached a good level of agreement (Cohen's Kappa .60, $p < .001$).

We tested the interaction effects between the use of emotion words and haiku topic valence on self-rated and other-rated happiness. As a measure of topic valence, we computed the average score across the two raters (as a continuous variable). In addition, we repeated the analyses using each rater's

valence score separately (using two dummies: positive vs. other and negative vs. other). Overall, we tested 30 different models (5 emotion word categories, 3 measures of haiku topic valence, and 2 dependent variables). For other-rated happiness, none of the interactions reached significance all ($ps > .05$). For self-rated happiness, the effect of anxiety and positive emotion words depended on haiku topic valence, yet the pattern of these interactions was not consistent with the idea that the use of positive emotion words is more diagnostic for negative valence texts and the use of negative emotion words is more diagnostic for positive valence texts (if anything, it was the opposite). The details regarding these analyses are provided in the Supplementary materials. Overall, as this study was not designed to test these hypotheses and the share of negative valence haikus was quite low, these results should be interpreted with caution.

Raters' judgment of haikus' happiness Raters were additionally asked to indicate to what extent the words "sad" and "cheerful" describe each haiku (recoded such that higher values indicate more happiness and averaged into a measure of *other-rated haiku happiness*, $r = .75, p < .001$). The association between average raters' judgment of targets' happiness and average raters' judgment of haikus' happiness was $r = .96 (p < .001)$, and the association between a rater's judgment of a specific haiku's happiness and this rater's judgment of this haiku's author's happiness was $r = .82 (p < .001)$. That is, raters assumed that happy people write happy haikus. We examined whether the same linguistic cues allow raters to judge some haikus as more and others as less happy and whether perceived haiku happiness shapes perceived targets' happiness. We conducted a serial mediation analysis where targets' self-rated happiness leads to the use of specific linguistic cues, which then shape raters' judgment of haikus' happiness, which ultimately affect raters' judgment of targets' happiness. Among the linguistic cues, we focused on the cues tested in the main analysis: sadness words and negative emotion words (with sadness words taken out of the score). We used multilevel regression and the Monte Carlo estimation method to construct the confidence intervals of the indirect effect. The indirect effect via "sadness words \rightarrow other-rated haiku happiness" was significant, .05 95% CI [.03; .08]; the indirect effect via "negative emotion words \rightarrow other-rated haiku happiness" was not, .005 95% CI [-.001; .013]. The path coefficients are presented in Figure S1 (Supplementary materials).

Discussion

Study 2 demonstrated that independent observers could judge targets' level of happiness at a better than chance level based on just three lines of their writing (fourteen words, on average). Like in Study 1, the use of certain linguistic cues

mediated the accuracy effect: unhappy targets wrote poems containing more words expressing negative emotions and sadness than their happy counterparts. These individual differences in word use were picked up by the raters and put to good use to make judgments of targets' happiness.

Use of negations: Additional analyses, Studies 1 and 2

An examination of the associations of self- and observer-rated well-being with all linguistic cues provided by LIWC (Table S1) showed that one linguistic cue was consistently associated with both self- and observer-rated well-being across the two studies: the use of negations. Specifically, both self- and other-rated SWB were associated with a spare use of negations (Study 1: $r = -.15, p < .030, r = -.30, p < .001$, respectively, Study 2: $r = -.19, p = .026, r = -.26, p = .002$, respectively). Therefore, in an additional set of analyses, we retested the mediation models with the use of negations as an additional parallel mediator, along with the use of negative emotion words. The results of these analyses are shown in Fig. 3. Indeed, the use of negations mediated the accuracy effect in both studies (Study 1: indirect effect 0.018, 95% CI [0.001; 0.04]; Study 2: indirect effect 0.02, 95% CI [0.003; 0.04]). At the same time, the mediation paths via anger and anxiety (Study 1) and sadness (Study 2) were unaffected by adding another mediator (Fig. 3).

Finally, we added the use of negation words to the serial mediation model in Study 2. Targets' self-rated happiness was modelled as leading to three linguistic cues (sadness words, negative emotion words and negations), the three linguistic cues led to raters' judgment of haikus' happiness, which led to raters' judgment of targets' happiness. These analyses also provided support for two indirect effects: via "sadness words -> haiku happiness" (.05 95% CI [.03; .08]) and via "negation words -> haiku happiness" (.02 95% CI [.01; .04]) (see Figure S2 for details).

Study 3: Pre-registered replication

While Studies 1 and 2 provided converging evidence of the accuracy effect in subjective well-being perception from texts, they have an important limitation. Both studies were part of larger research projects and included more measures than the ones we focused on in the present paper (subjective well-being) and the hypotheses of neither of these studies were preregistered. Therefore, following the request of one of the anonymous reviewers, we conducted a direct replication of one of the studies. As travel reports have a higher ecological validity (that is, in real life, people are more likely to write travel reports than haikus), for the replication, we selected Study 1. Compared to Study 1, Study 3 had the

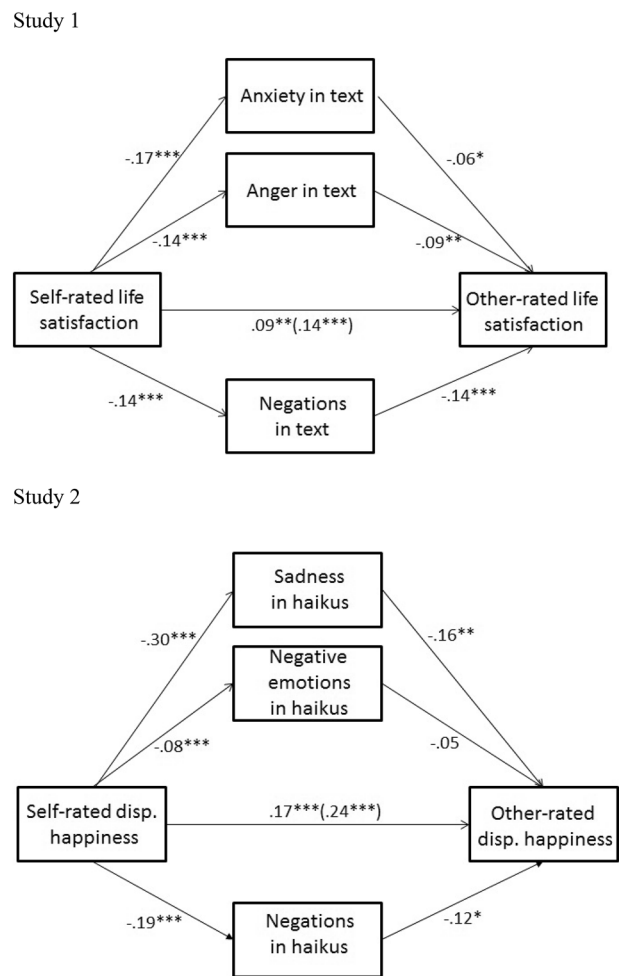


Fig. 3 Mediation analysis via emotion and negation words, Studies 1 and 2

following two modifications. First, we only included the measures we were interested in—self-rated and other-rated life satisfaction. Second, to make sure that reminding participants of their holiday time does not affect their life satisfaction judgment, targets reported their life satisfaction before completing the writing assignment. Hypotheses, measures, data collection, and analyses were preregistered (<http://aspredicted.org/blind.php?x=y3pa4x>). We did not deviate from the preregistered plan in any way, except for ending up with 12 targets more than planned (see below for details). Individuals who participated in Study 1 or 2 as either targets or raters were not eligible for this study.

Method

Targets

Following our preregistered plan, we recruited 200 individuals from MTurk. 19 participants did not provide travel

reports (e.g., some of them copied unrelated random texts from some websites; other wrote something unrelated, like “ha ha no way”). To compensate for these lost cases, we recruited an additional 30 participants (we oversampled as we were concerned that some of these newly recruited participants will provide unusable texts too). At the end, we had 212 targets who provided usable travel reports (mean age = 37.23, 44.3% female).

At the beginning of the survey, targets reported their overall life satisfaction. We used the same measure as in Study 1: participants rated their life overall on a scale ranging from 0 = “the worst possible life overall” to 10 = “the best possible life overall” (Cantril 1965). Afterwards, participants were asked to complete the same writing task as in Study 1. The instructions did not deviate from Study 1 in any way. On average, participants wrote 76.88 words ($SD = 41.77$). At the end, participants indicated their gender and age.

Raters

220 (mean age = 37.01, 42.7% female) individuals were recruited from MTurk to rate the obtained texts (MTurkers who participated as targets were not eligible). On average, raters rated five texts and texts were rated by five raters. Raters assessed each target’s life satisfaction using the same scale as the targets did (0 = “the worst possible life overall” to 10 = “the best possible life overall”). This was the only measure that the participants rated the texts on. At the end, the participants indicated their gender and age.

Results

Accuracy

Replicating the results of Study 1, the correlation between average observer ratings of targets’ life satisfaction and targets’ self-ratings was $r = .19$ ($p = .006$). The multilevel regression analysis supported this finding: self-rated life satisfaction was positively related to other-rated life satisfaction ($b = .10$, $p = .011$). Similar to Study 1, raters on average overestimated targets’ happiness by .72 on an 11-point scale (raters: $M = 8.02$, $SD = 1.24$; targets: $M = 7.30$, $SD = 2.23$; $t(211) = 2.67$, $p < .001$).

Linguistic analyses: Cue validity

Of all emotion word categories, only the use of positive emotion words was related to self-rated life satisfaction ($r = .14$, $p = .050$; all other $ps > .09$; Table 1).

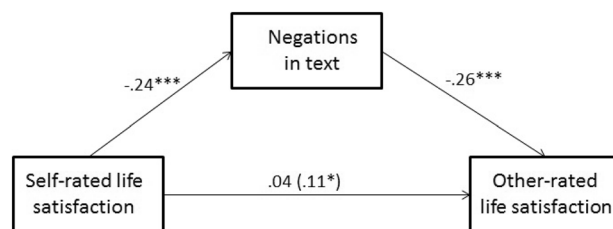


Fig. 4 Mediation analysis via negation words, Study 3

Linguistic analyses: Cue utilization

Like in Study 1, other-rated life satisfaction was positively associated with the use of negative emotion words ($r = -.25$, $p < .001$), anxiety words ($r = -.15$, $p = .032$) and anger words ($r = -.14$, $p = .039$), but not positive emotion words ($r = -.002$, $p = .98$) and sadness words ($r = -.13$, $p = .069$; Table 1).

Overall, comparing cue validity and cue utilization (Table 1) shows that the emotion words that the raters relied on to judge targets’ life satisfaction do a poor job predicting targets’ self-rated life satisfaction. Hence, emotion words are unlikely to mediate the association between other- and self-rated life satisfaction and cannot explain the accuracy in life satisfaction perception from texts in this study.

Exploratory analysis: Use of negations

As the use of negations represented a significant (although not predicted) mediator of the accuracy effect in both Study 1 and Study 2, we explored it in Study 3 too. Like in Studies 1 and 2, targets who were more satisfied with life were less likely to use negations in their texts ($r = -.24$, $p < .001$). Targets’ use of negations was negatively associated with raters’ judgment of targets’ life satisfaction ($r = -.44$, $p < .001$; Table S1).

Next, we examine targets’ use of negation words as a mediator of the association between other-rated and self-rated life satisfaction. Like in Studies 1 and 2, we used multilevel regression and the Monte Carlo estimation method to construct the confidence intervals of the indirect effect. The indirect effect was significant: .062, 95% CI [.040; .088]. The path coefficients are presented in Fig. 4. Targets’ self-rated life satisfaction was negatively related to their use of negations ($b = -.24$, $p < .001$) and the use of negations ($b = -.26$, $p < .001$) in turn predicted observer ratings of targets’ life satisfaction. Accounting for targets’ use of negations rendered the association between other- and self-rated life satisfaction non-significant ($b = .11$, $p = .011$ vs. $b = .04$, $p = .30$).

Discussion

The pre-registered Study 3 replicated the main finding of Studies 1 and 2: individuals could accurately judge others' life satisfaction from texts at a better than chance level. However, in contrast to Studies 1 and 2, the use of emotion words did not mediate this effect. Even though observers' estimates of targets' life satisfaction were based on targets' use of negative emotion words, negative emotion words turned out to be an invalid cue to targets' life satisfaction in this study. Instead, like in Studies 1 and 2, targets' use of negations fully explained the accuracy effect: more satisfied targets were less likely to use negations in their texts and observers picked this up and (successfully) used this cue in their judgment of targets' life satisfaction.

General discussion

In a draft of his Nobel Prize speech, Ernest Hemingway wrote that if the writer “has written well, everything that is him has gone into the writing” (“Ernest Hemingway: A Storyteller's Legacy” 2017). Indeed, writing can contain multiple cues to individuals' age, gender, personality and even psychological health (Pennebaker et al. 2003; Schwartz et al. 2013). How good are people in making use of such cues and how accurately can people judge others' happiness based on the texts they write?

The present studies showed that people can judge unacquainted others' level of subjective well-being, based on brief excerpts of the texts they wrote, at a better than chance level. How does the degree of self-other agreement in our studies fare relative to the effect sizes reported in the literature on well-being judgment among dating couples and friends? According to meta-analytic findings (Schneider and Schimmack 2009), an average self-other agreement in these studies is $r = .42$ (credibility interval: .39–.45). Although this is higher than the effects reported here ($r = .22$ in Study 1, $r = .37$ in Study 2, and $r = .19$ in Study 3) and people seem to be worse at accurately judging a complete stranger's SWB than the SWB of their friends and dating partners, we consider the level of accuracy achieved in the present studies quite impressive given that the basis for the judgment was just a few lines of text.

What allowed this degree of accuracy? We expected targets' use of emotion words in texts to provide raters with cues, making an accurate judgment of targets' happiness possible. Indeed, in the first two studies, targets' self-reports of SWB were reflected in the use of negative emotion words in the texts they wrote and raters, in turn, were able to pick up on these linguistic cues and, consequently, judged targets' SWB at a better than chance level. However, the specific emotion word categories that served as valid cues

to well-being differed across these studies, with anger and anxiety words being valid cues in Study 1 and sadness words being valid cues in Study 2. Even more importantly, a pre-registered replication (Study 3) failed to provide support to the use of negative emotion words as valid cues of life satisfaction: even though raters used negative emotion words to judge targets' life satisfaction, none of the negative emotion word categories was significantly associated with targets' self-reported life satisfaction. The associations between self-reports of SWB and the use of positive emotion words didn't show a consistent pattern across the studies either.

The fact that we didn't detect consistent associations between self-rated SWB and the use of emotion words in texts might have important methodological implications. Using individual differences in emotional language as an indicator of SWB is getting increasingly common in social sciences. For example, the frequency of positive emotional language in tweets was recently used to draw conclusions about the effect of political ideology (Wojcik et al. 2015) and religiosity (Ritter et al. 2014) on twitters' happiness. Our results suggest that, although the analysis of linguistic cues in texts might provide insights into individuals' well-being, the validity of these measures is not ironclad and potentially depend on the specific emotion word category, the type of writing used and potentially further factors. Hence, relying on the use of emotion words as indicators of subjective well-being might be premature (also see Luhmann 2017) and more studies are needed to determine the use of what emotion words in what types of texts represent valid cues to text authors' subjective well-being.

If the use of emotion words does not consistently account for the accuracy effect, what linguistic cues do? Exploratory analyses showed that one linguistic cue was consistently associated with both self-rated and other-rated subjective well-being in all three studies: the use of negations. More (vs. less) happy and satisfied targets were less likely to use negations in their texts and raters picked up on this cue, which ultimately allowed them to make accurate judgments of targets' life satisfaction. The mediation analyses confirmed that the use of negations explained a significant portion of the accuracy effect in all three studies. The use of negations has been rarely the focus of attention in studies on the linguistic expression of well-being. Previous studies that considered this linguistic marker provided mixed findings. For example, negations have been shown to be especially frequent in individuals' recollections of negative life events (e.g., loss and trauma) (Hargitai et al. 2007) and particularly often used by individuals with a dismissing attachment style (Cassidy et al. 2012). Yet, a recent analysis of everyday spoken language reported the use of negations to be unrelated to self-reported positive or negative emotions (Sun et al. 2019). Overall, the fact that the use of negations consistently emerged as a mediator of the accuracy effect in the

present three studies suggests that negations might represent an understudied linguistic marker of well-being. Hence, more studies are needed to understand why individuals with lower well-being might be more likely to use negations and why observers believe that the use of negations is a valid indicator of poor well-being.

The present research has several limitations. Although single-item measures of life satisfaction have been validated in previous research (Cheung and Lucas 2014), using a single-item measure in Study 1 (and Study 3) could have contributed to a weaker accuracy effect (compared to, for example, Study 2) (Schneider and Schimmack 2009). Also, while the present studies provided evidence for judgment accuracy of SWB from texts of a relatively private nature, it is important to examine whether this accuracy effect would replicate with respect to publicly shared texts, such as blogs, tweets or status updates in social networks. We speculate that increased impression management concerns common on social network sites might reduce individuals' use of linguistic cues indicative of their actual SWB level and make the task of accurate SWB judgment even harder and the perception accuracy substantially smaller. Similarly, while we have shown that people might be able to accurately judge others' SWB from samples of creative writing (haikus), this accuracy effect might be lower (or even absent) in professional creative writers, as their goal is often not to express themselves but to elicit emotions (and sometimes behaviors) in their readers. Finally, cultures differs with respect to emotional display rules (Matsumoto et al. 2008), which could undermine the accuracy of well-being detection from texts. Given the rising internalization, increasing migration and intercultural contacts, another interesting question for future studies is whether a similar accuracy level can be achieved in the context of intercultural communication.

To conclude, the present research suggests that not only great literary works might reflect the authors' inner states, as Ernest Hemingway pointed in a draft of his Nobel Prize speech, but everyday texts of ordinary people might also contain several cues that, when used well, can give accurate insights into their well-being.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Bazarova, N. N., Taft, J. G., Choi, Y. H., & Cosley, D. (2013). Managing impressions and relationships on Facebook: Self-presentational and relational concerns revealed through the analysis of language style. *Journal of Language and Social Psychology*, 32(2), 121–141.
- Brunswick, E. (1956). *Perception and the representative design of experiments*. Berkeley, CA: University of California Press.
- Cantril, H. (1965). *The pattern of human concerns*. New Brunswick, NJ: Rutgers University Press.
- Cassidy, J., Sherman, L. J., & Jones, J. D. (2012). What's in a word? Linguistic characteristics of Adult Attachment Interviews. *Attachment and Human Development*, 14(1), 11–32.
- Cheung, F., & Lucas, R. (2014). Assessing the validity of single-item life satisfaction measures: Results from three large samples. *Quality of Life Research*, 23, 2809–2818.
- Ernest Hemingway: A Storyteller's Legacy. (2017). John F. Kennedy. Presidential Library and Museum. Retrieved February 22, 2017, from <https://www.jfklibrary.org/Research/The-Ernest-Hemingway-Collection/Online-Resources/Storytellers-Legacy.aspx?p=13>.
- Fernandez, K. C., Levinson, C. A., & Rodebaugh, T. L. (2012). Profiling: Predicting social anxiety from Facebook profiles. *Social Psychological and Personality Science*, 3(6), 706–713.
- Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, 11(2), 77–83.
- Hargitai, R., Naszódi, M., Kis, B., Nagy, L., Bóna, A., & László, J. (2007). Linguistic markers of depressive dynamics in self-narratives: Negation and self-reference. *Empirical Text and Culture Research*, 3, 26–38.
- Holleran, S. E., & Mehl, M. R. (2008). Let me read your mind: Personality judgments based on a person's natural stream of thought. *Journal of Research in Personality*, 42, 747–754.
- Hox, J. (2002). *Multilevel analysis: Techniques and applications*. New Jersey: Lawrence Erlbaum Associates.
- Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, 103, 54–69.
- Küfner, A. C. P., Back, M. D., Nestler, S., & Egloff, B. (2010). Tell me a story and I will tell you who you are! Lens model analyses of personality and creative writing. *Journal of Research in Personality*, 44, 427–435.
- Kuppens, P., Realo, A., & Diener, E. (2008). The role of positive and negative emotions in life satisfaction judgment across nations. *Journal of Personality and Social Psychology*, 95, 66–75.
- Lee, C. H., Kim, K., Seo, Y. S., & Chung, C. K. (2007). The relations between personality and language use. *Journal of General Psychology*, 134, 405–413.
- Li, J., & Chignell, M. (2010). Birds of a feather: How personality influences blog writing and reading. *International Journal of Human-Computer Studies*, 68, 589–602.
- Lin, W.-F., Lin, Y.-C., Huang, C.-L., & Chen, L. H. (2016). We can make it better: "We" moderates the relationship between a compromising style in interpersonal conflict and well-being. *Journal of Happiness Studies*, 17(1), 41–57.
- Lin, H., Tov, W., & Qiu, L. (2014). Emotional disclosure on social networking sites: The role of network structure and psychological needs. *Computers in Human Behavior*, 41, 342–350.

- Liu, P., Tov, W., Kosinski, M., Stillwell, D. J., & Qiu, L. (2015). Do Facebook status updates reflect subjective well-being? *Cyberpsychology, Behavior, and Social Networking*, *18*(7), 373–379.
- Luhmann, M. (2017). Using big data to study subjective well-being. *Current Opinion in Behavioral Sciences*, *18*, 28–33.
- Matsumoto, D., Yoo, S. H., Fontaine, J., Anguas-Wong, A. M., Arriola, M., Ataca, B., et al. (2008). Mapping expressive differences around the world: The relationship between emotional display rules and individualism versus collectivism. *Journal of Cross-Cultural Psychology*, *39*(1), 55–74.
- Mehl, M. R. (2006a). The lay assessment of subclinical depression in daily life. *Psychological Assessment*, *18*, 340–345.
- Mehl, M. R. (2006b). Quantitative text analysis. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology*. Washington, DC: American Psychological Association.
- Nestler, S., & Back, M. D. (2013). Applications and extensions of the lens model to understand interpersonal judgments at zero acquaintance. *Current Directions in Psychological Science*, *22*(5), 374–379.
- Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2007). *Linguistic inquiry and word count (LIWC): LIWC 2007*. Mahwah, NJ: Erlbaum.
- Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, *77*, 1296–1312.
- Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, *54*, 547–577.
- Qiu, L., Lin, H., Leung, A. K., & Tov, W. (2012a). Putting their best foot forward: Emotional disclosure on Facebook. *Cyberpsychology, Behavior and Social Networking*, *15*(10), 569–572.
- Qiu, L., Lin, H., Ramsay, J., & Yang, F. (2012b). You are what you tweet: Personality expression and perception on twitter. *Journal of Research in Personality*, *46*, 710–718.
- Ritter, R. S., Preston, J. L., & Hernandez, I. (2014). Happy Tweets: Christians are happier, more socially connected, and less analytical than atheists on Twitter. *Social Psychological and Personality Science*, *5*, 243–249.
- Rodriguez, A. J., Holleran, S. E., & Mehl, M. R. (2010). Reading between the lines: The lay assessment of subclinical depression from written self-descriptions. *Journal of Personality*, *78*, 575–598.
- Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology*, *9*(4), 283–294.
- Rude, S. S., Gortner, E.-M., & Pennebaker, J. W. (2004). Language use of depressed and depression-vulnerable college students. *Cognition and Emotion*, *18*, 1121–1133.
- Schneider, L., & Schimmack, U. (2009). Self-informant agreement in well-being ratings: A meta-analysis. *Social Indicators Research*, *94*, 363–376.
- Schneider, L., & Schimmack, U. (2010). Examining sources of self-informant agreement in life-satisfaction judgments. *Journal of Research in Personality*, *44*, 207–212.
- Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., et al. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS ONE*, *8*, e73791.
- Selig, J. P., & Preacher, K. J. (2008). Monte Carlo method for assessing mediation: An interactive tool for creating confidence intervals for indirect effects [Computer software]. Retrieved from <http://quantpsy.org/>.
- Settanni, M., & Marengo, D. (2015). Sharing feelings online: Studying emotional well-being via automated text analysis of Facebook posts. *Frontiers in Psychology*, *6*, 1045.
- Stirman, S. W., & Pennebaker, J. W. (2001). Word use in the poetry of suicidal and nonsuicidal poets. *Psychosomatic Medicine*, *63*, 517–522.
- Sun, J., Schwartz, H. A., Son, Y., Kern, M. L., & Vazire, S. (2019). The language of well-being: Tracking fluctuations in emotion experience through everyday speech. *Journal of Personality and Social Psychology*.
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*, 24–54.
- Tov, W., Ng, K. L., Lin, H., & Qiu, L. (2013). Detecting well-being via computerized content analysis of brief diary entries. *Psychological Assessment*, *25*, 1069–1078.
- Tskhay, K. O., & Rule, N. O. (2014). Perceptions of personality in text-based media and OSN: A meta-analysis. *Journal of Research in Personality*, *49*, 25–30.
- Veenhoven, R. (1989). *How harmful is happiness? Consequences of enjoying life or not*. Rotterdam: University of Rotterdam.
- Watson, D., Hubbard, B., & Wiese, D. (2000). Self–other agreement in personality and affectivity: The role of acquaintanceship, trait visibility, and assumed similarity. *Journal of Personality and Social Psychology*, *78*, 546–558.
- Wojcik, S. P., Hovasapian, A., Graham, J., Motyl, M., & Ditto, P. H. (2015). Conservatives report, but liberals display, greater happiness. *Science*, *347*, 1243–1246.
- Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality*, *44*, 363–373.
- Yeagley, E., Morling, B., & Nelson, M. (2007). Nonverbal zero-acquaintance accuracy of self-esteem, social dominance orientation, and satisfaction with life. *Journal of Research in Personality*, *41*, 1099–1106.
- Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing multilevel mediation using hierarchical linear models: Problems and solutions. *Organizational Research Methods*, *12*, 695–719.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.