
Works Cited

Yaden, David B., et al. "Characterizing Empathy and Compassion Using Computational Linguistic Analysis." *Emotion*, vol. 24, no. 1, Feb. 2024, pp. 106–15. *EBSCOhost*, https://doi.org/10.1037/emo0001205.supp (Supplemental).

Characterizing Empathy and Compassion Using Computational Linguistic Analysis

By: David B. Yaden
Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine;
Department of Psychology, University of Pennsylvania;
Salvatore Giorgi
Department of Computer and Information Science, University of Pennsylvania
Matthew Jordan
Department of Psychology, Yale University
Anneke Buffone
Department of Psychology, University of Pennsylvania
Johannes C. Eichstaedt
Department of Psychology and Institute for Human-Centered AI, Stanford University
H. Andrew Schwartz
Department of Computer Science, Stony Brook University
Lyle Ungar
Department of Computer and Information Science, University of Pennsylvania
Paul Bloom
Department of Psychology, Yale University

Acknowledgement: This research was supported in part by a grant from the Templeton Religion Trust. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Templeton Religion Trust.

David B. Yaden and Salvatore Giorgi are equal contributions.

David B. Yaden served as lead for writing–original draft and writing–review and editing, contributed equally to supervision, and served in a supporting role for conceptualization and investigation. Salvatore Giorgi served as lead for data curation, formal analysis, methodology, and writing–review and editing. Matthew Jordan contributed equally to methodology. Anneke Buffone contributed equally to writing–original draft and served in

a supporting role for writing–review and editing. Johannes C. Eichstaedt served in a supporting role for conceptualization and writing–review and editing. H. Andrew Schwartz served in a supporting role for conceptualization, project administration, and writing-review and editing. Lyle Ungar served as lead for funding acquisition and served in a supporting role for conceptualization, data curation, formal analysis, writing–review and editing. Paul Bloom served in a supporting role for conceptualization, investigation, project administration, and resources. David B. Yaden and Lyle Ungar contributed to project administration equally. Matthew Jordan and Paul Bloom contributed to writing–review and editing equally. Matthew Jordan and Anneke Buffone contributed to conceptualization equally.

Empathy for suffering, vulnerable, or oppressed others has been argued to be central to moral character and collective well-being. This view has been articulated, in somewhat different forms by Smith (1759/1976) and Hume (1739/1978), and has formed the basis for both theories of the origin of morality in the species (De Waal, 2010) and theories of moral development in children (Batson, 2011; Bloom, 2017; Hoffman, 2008, for discussion). The core argument here is that only by feeling the suffering and pleasures of others do we care about reducing the former and increasing the latter. Under some versions of this claim, this occurs, at least in part, through vicarious suffering—empathy causes people to feel the pain of others, and, to reduce their own suffering, they act to make other's pain go away (see Baron-Cohen & Wheelwright, 2004; Batson, 2010; Hoffman, 2008; Toi & Batson, 1982; Zaki, 2016).

Others have argued that empathic distress is insufficient as a foundation for morality (e.g., Bloom, 2017; Prinz, 2011). Instead, altruism is motivated by more general feelings of compassion—caring for others without necessarily mirroring their negative feelings. There is ongoing debate as to the precise relationship between empathy and compassion and their relative contributions to moral and social behavior (Batson, 2010; Batson et al., 1987; Buffone et al., 2017; Decety & Cowell, 2014; De Waal, 2010; Jordan et al., 2016; Morelli et al., 2015).

One proposal is that the difference between the two is as follows. Empathy is a self-focused process in which the target's states are internalized and then become the focus of the empathizer's attention. This involves emotional and sometimes behavioral matching, a tendency to experience others' emotions and thoughts as one's own (Singer & Klimecki, 2014). Evidence suggests that such matching can be relatively automatic and is more likely to occur among those high on trait empathy (Buffone et al., 2017; Decety & Jackson, 2006; Singer & Klimecki, 2014). Additionally, these self-focused traits and states are linked, in some populations, to negative health outcomes and burnout (Hojat et al., 2002).

In contrast, compassion is an other-focused process in which the mental and physical state of the target remains front and center in the mind of the compassionate individual. This leads to increased attention to others and an intention to assist (Bloom, 2017; Davis, 1983; Jordan et al., 2016). Those predisposed to other-focus tend to care more about others' pain and suffering without feeling it themselves, and this trait is related to better health and approach motivation (Buffone et al, 2017, Singer & Klimecki, 2014). Other-focused processes are thought to be more positive in affect (e.g., feelings of tenderness, concern for well-being, warmth). Additionally, these other-focused states may be subject to greater cognitive control and may be associated with invigorating arousal, which is ideal for goal pursuit (Buffone et al., 2017).

In support of this analysis, Jordan et al. (2016) found that their measure of empathy, including subscales for empathy and behavioral contagion, correlated highly with the personal distress sub-scale of the Interpersonal Reactivity Index (IRI; Davis, 1983), and people high on that factor are both less prosocial overall and become less prosocial in the face of negative emotions in the target of their empathy. On the other hand, subscales of

empathic concern (what we would label compassion) are correlated with perspective taking and people high on that factor are both more prosocial and become more prosocial in the presence of a target in distress.

While empathy and compassion have been examined using neuroimaging and various psychometric approaches, one particularly promising additional method that might provide further insight is computational linguistic analysis. The computational linguistic analysis identifies linguistic features (words, phrases) that are associated with a given outcome of interest. Such language-based assessments, typically using large social media datasets, yield insights into personality, emotions, experiences, behaviors, and demographic characteristics (Curtis et al., 2018; Kern et al., 2014; Park et al., 2015; Schwartz et al., 2013b; Yaden et al., 2016, 2018, 2021), mental health (Eichstaedt et al., 2018; Preoţiuc-Pietro et al., 2015), well-being (Schwartz et al., 2013a, 2016) and physical health (Eichstaedt et al., 2015) of users. Linguistic analysis can also provide insight into the emotional, social, and behavioral differences between those with different traits. For example, Schwartz et al., (2013b) found strong associations between language use and personality: those high in openness to experience used words like “dream”; conscientiousness, “ready for”; extroversion, “party”; agreeable, “thank you”; and neuroticism, “sick of.” Findings like these can help to better categorize the constructs in question and, in some cases, can lead to hypothesis generation regarding the implication of psychological processes. Beyond characterization, identifying the linguistic correlates associated with constructs of interest can allow future studies to predict these constructs on the basis of language alone. Thus, identifying linguistic correlates of a psychological construct can lead to models that can unobtrusively predict levels of that trait without the need for self-report surveys if sufficient language is available (Eichstaedt et al., 2021).

In this research, we apply computational linguistic analysis to autobiographic posts shared on Facebook to understand which linguistic tendencies predict empathy and compassion at the individual level. Critically, this approach will allow us to examine whether self-focus and other-focus track with empathy and compassion in an externally valid setting. By mapping linguistic tendencies onto empathy and compassion, as well as several measures of life outcomes (health, well-being, prosocial behaviors), we explore a set of traits and outcomes that are central to social and moral psychology.

Method

Participants and Procedures

Study participants were recruited via Qualtrics paneling to an online Qualtrics survey including items on empathy, health, well-being, and prosocial behavior . After providing informed consent, participants were asked to share their own Facebook statuses across the past four years and were then redirected to a survey. The IRB at the University of Pennsylvania approved this study (protocol 816091). Out of 5,542 consenting respondents, a total of 2,931 completed the survey (i.e., all empathy, compassion, and perceived stress items as well as age, gender, income, and education) without failing any attention checks (see the online supplemental materials for a full break down of participant dropout). Only participants who had written 500 or more words in their status updates on Facebook were retained, as this has been established as a useful threshold for computational linguistic analysis (Kern et al., 2016). Additionally, participants must have posted at least five Facebook status updates within the previous 180 days to ensure they are active users on the platform. The final sample (N = 2,781, 58.8% female, *Mdn*age = 36 years, *SD* = 12.5 years, range = 18–85) collectively shared 2,356,916 Facebook status updates (median status updates: 482; mean status updates: 847; *SD* = 1,064).

In order to replicate our findings (i.e., the factor structure of the empathy and compassion scales as well as the language analysis), we collected a second set of 844 participants using Qualtrics panels. Again, we limited our

study to participants who provided consent, self-reported age and gender, and who answered each item in our empathy scales. In the end, this second sample consisted of 794 participants (71.5% female; *Mdn*age = 42 years, *SD* = 12.0 years, range = 18–76). We used this sample to validate our factor analysis of the compassion and empathy scales.

Finally, we examined the relationships between empathy and compassion with personality. Using the MyPersonality dataset (Kosinski et al., 2015), we examined how the language of high compassion (controlling for empathy) and the language of high empathy (controlling for compassion) were distributed along personality traits. We took 16,507 users from MyPersonality who shared Facebook data, wrote 1,000 words or more in their status updates, and completed a Big-Five personality questionnaire based on the International Personality Item Pool proxy for the NEO Personality Inventory (Costa & McCrae, 2008). While a 500-word minimum was used in the other samples, a 1,000-word minimum was chosen here to match previous work on personality which used this data set (Schwartz et al., 2013b).

Individual-Level Self-Report Measures

Empathy and Compassion Items

Compassion items were comprised of items from the empathic concern subscale of the IRI (Davis, 1983) along with items generated to measure imagine-other perspective taking (Buffone et al., 2017) which focuses on taking on other's thoughts and feelings without making their suffering one's own. Empathy items were comprised of items from the Empathy Index (Jordan et al., 2016) as well as items generated to measure lack of control (automatic contagion), making other's suffering one's own, imagine-self perspective taking (Buffone et al., 2017), and a further behavioral aspect of crying with a suffering person.

Socioeconomics

The survey included measures of income and education.

Health

Participants were asked various health-related items and scales: Cohen's 10-item stress scale (Cohen et al., 1983), mental health, physical health, missing work for health reasons, and weight.

Lifestyle

Our lifestyle measures included exercise frequency and excessive drinking.

Prosocial Behaviors

Prosocial behaviors were assessed through self-report of donations and volunteering.

Analytic Approach

We performed a factor analysis on the compassion and empathy items. Because our primary aim was to characterize and explore compassion and empathy, and because we expected these factors to correlate with one another, we extracted two factors using an oblique (promax) rotation.

Language Correlates of Empathy

We employed both "top-down" and "bottom-up" computational linguistic analysis approaches (Schwartz et al., 2013b). The top-down approach is the Linguistic Inquiry Word Count (LIWC) program, which includes a set of

theoretical, researcher-created language categories (e.g., the word “happy” is counted toward the *Positive Emotion* category; Pennebaker et al., 2015). We compared empathy and compassion across *all* LIWC categories. The bottom-up approach uses differential language analysis (DLA; Schwartz et al., 2013b), which utilizes topics, or groups of semantically related words and phrases that have been algorithmically clustered using Latent Dirichlet allocation (LDA; Blei et al., 2003). The open-source Python package Differential Analysis ToolKit (Schwartz et al., 2017) was used for all linguistic analyses. Both the top-down and bottom-up approaches occur in three stages: linguistic feature extraction, correlational analysis, and visualization (Kern et al., 2016; Schwartz et al., 2013b).

Linguistic Feature Extraction

Given each user’s Facebook language, we extract two types of features: (a) *words* and *word phrases* and (b) *topics*. Facebook statuses are broken up into *words* (or tokens) using a publicly available tokenizer designed to capture standard language as well as language specific to social media (e.g., misspellings, emoticons, and hashtags). Sequences of two- and three-word *phrases* were kept if they are more likely to appear together than by chance (Kern et al., 2016). Each *word* and *word phrase* was retained if used by at least 10% of the participants, in order to remove rare language and minimize spurious correlations. Each feature was encoded as the relative frequency of that feature per user: the number of times the *word* or *word phrase* was used divided by the total number of *words* and *word phrases*.

For the *topic* features we used a set of 2,000 social media-based topics derived from approximately 15 million posts from the MyPersonality Facebook data set (Schwartz et al., 2013b). These topics were automatically derived using LDA (Blei et al., 2003). LDA uses a generative model which assumes each document (or Facebook post) contains a distribution of topics, which is assumed to be a distribution of words. The output of the LDA algorithm is a set of semantically grouped words with weights for each word in the topic. Highly weighted words within the same topic typically appear in similar contexts. Topics were created using the Mallet software package (McCallum, 2002) which uses Gibbs sampling (Gelfand & Smith, 1990) to estimate the latent variables of the topic. All default Mallet settings were used except *alpha*, a prior on the expected topics per document, which was set to 3. The setting was chosen under the assumption that Facebook posts contain fewer topics than standard text (such as newspaper articles) where LDA is traditionally applied. This set of 2,000 topics has been used to capture linguistic variation in satisfaction with life (Schwartz et al., 2013a), personality (Park et al., 2015), and heart disease (Eichstaedt et al., 2015).

Correlation Analysis

Each linguistic feature is then standardized and fed into an ordinary least-squares linear regression as our independent variable, with compassion and empathy measures as our dependent variables. When looking at either compassion or empathy, we include the other as a covariate in the model. To help the reader, the results of these analyses are called empathy without compassion (i.e., empathy controlled for compassion) and compassion without empathy (i.e., compassion controlled for empathy). Thus, all analyses use multiple regression (since either empathy or compassion is included as a covariate), but we never include two or more linguistic variables within the same regression. Since this produces tens of thousands of correlations, we correct for multiple comparisons using a Benjamini-Hochberg false discovery rate correction (Benjamini & Hochberg, 1995). After correction, correlations are significant if they have a two-tailed *p*-value less than $\alpha = .05$.

Visualization

To visualize the resulting correlations all *words*, *word phrases,* and *topics* are visualized via word clouds. Words and word phrases are grouped together into a single word cloud where the color of the word indicates its frequency of use and the size indicates the strength of the correlation with empathy (standardized βs from the multiple linear regression). Larger words have larger correlations while grey, blue, and red words indicate less frequent, moderately frequent, and highly frequent words, respectively. Topic clouds show the top 15 most prevalent words within the topic and are sized according to their posterior likelihood (how often they appear as a representative of the topic).

Personality and the Language of Empathy and Compassion

For each of the 16,507 users in the MyPersonality data set, we extracted 2,000 topic features (identical to the topics described above) and subsetted the top 25 most positively correlated topics with both compassion controlling for empathy and empathy controlling for compassion (50 topics total). We then took each user's five-factor model personality factor scores (openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism) and calculated the Pearson *r* for each of the 50 empathy and compassion language topics. Thus, this analysis examines survey-based personality and its relationship to the language associated with empathy and compassion.

Health Correlates

We also looked at various categories of survey responses (demographics, socioeconomics, health, lifestyle, and prosocial behaviors) to see how well empathy controlling for compassion and compassion controlling for empathy predict these outcomes. For each survey item, we standardize compassion and empathy (independent variables) and outcome survey items (dependent variables) and compute an ordinary least-squares linear regression. Again, when looking at compassion or empathy, we include the other construct as a covariate in the model.

Results

Factor Analysis

Factor extraction criteria suggested two factors that were in line with the hypothesized compassion/empathy distinction (Figure 1). Three items were removed from the compassion scale due to negative loading, thus leaving seven total items. We removed the item with the lowest loading in the empathy scale to minimize participant burden (i.e., shorten the scale for future use) as well as keep the same number of items in both the compassion and empathy scales. This overall structure is held in both the current dataset (see Table 1) and the follow-up dataset (see Table S2 in the online supplemental materials). In the current dataset, with a promax rotation (see Table S3 in the online supplemental materials for varimax rotation loadings), 24% of the variance is explained by the empathy factor (loadings 0.59–0.82) and 22% by the compassion factor (loadings 0.51–0.80).

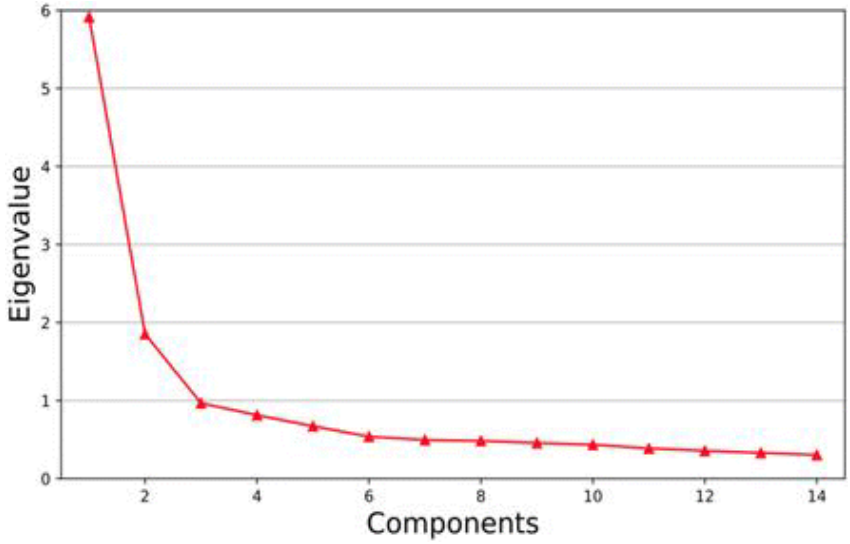

Figure 1.

Table 1

Scale Questions With Promax Rotated Factor Loadings

Question	Construct	Original scale	Scale	Promax Factor 1	Promax Factor 2
If I hear a story in which someone is scared, I will imagine how scared I would be in that situation.	Empathy	Jordan	1–7	0.59	0.07
I sometimes find myself feeling the emotions of the people around me, even if I do not try to feel what they are feeling.				0.66	0.08
If I see someone fidgeting, I will start feeling anxious too.				0.61	−0.13
I tend to make other people's suffering my own. That is, I take on other people's sadness and upset when faced with their suffering.	Empathy	ISPT	1–9	0.78	−0.03
Other people's sadness or upset is contagious for me.				0.82	−0.04
When I see someone cry, I am very likely to cry with them.				0.66	0.04
If I hear about someone very similar to me experiencing a tragedy, I automatically experience their sadness and suffering as my own.				0.67	0.14
I often have tender, concerned feelings for people less fortunate than me.	Compassion	IRI—Empathic Concern	1–5	−0.06	0.66
When I see someone being taken advantage of, I feel kind of protective toward them.				−0.11	0.63
I am often quite touched by things that I see happen.				0.09	0.56
I would describe myself as a pretty soft-hearted person.				0.09	0.51
When somebody tells me about a problem they are facing, I try to imagine how this person must feel with regard to their situation.	Compassion	IOPT	1–9	−0.02	0.80
I often try to imagine how another person must feel with regard to what happened to them.				0.00	0.77
When I hear about a terrible event that happened to someone else (e.g., in conversation, on the news, etc.) I immediately try to imagine how those affected must feel.				0.14	0.67

Note. IOPT = imagine-other perspective taking; ISPT = imagine-self perspective taking; IRI = Interpersonal Reactivity Index.

Scale Questions With Promax Rotated Factor Loadings

Using the items in Table 1 to create both empathy and compassion scales resulted in a correlation of $r = .53$, $p < .001$ CI = [0.50, 0.56], between the two constructs.

Language Results

Table 2 shows the results of correlating LIWC with compassion and empathy in our main sample. As opposed to all of the other analyses in the manuscript, these correlations are not controlling for the opposite construct. Out of the top 10 most correlated LIWC categories, five categories are correlated with both compassion and empathy. To further emphasize the shared language between the constructs, we compared all LIWC category correlations across both empathy and compassion. To do this, we correlated each of the 73 LIWC categories with both empathy and compassion. We then created vectors for both compassion and empathy of the

resulting Pearson *r* values (i.e., two vectors of 73 correlations). We then correlated those two vectors, which resulted in a Pearson's *r* of .98.

Table 2

Top 10 Positively Correlated LIWC Categories With Empathy (First 10 Lines) and Compassion (Last 10 Lines)

		Empathy		Compassion	
Category	Words	β	95% CI	β	95% CI
Ppron	I, my, you, me, your	.18***	[0.14, 0.21]	.15***	[0.11, 0.19]
Affect	Love, good, happy, :), great	.17***	[0.13, 0.20]	.15***	[0.12, 0.19]
I	I, my, me, i'm im	.17***	[0.13, 0.20]	.10***	[0.06, 0.13]
Pronoun	I, my, you, it, this	.15***	[0.12, 0.19]	.13***	[0.09, 0.16]
Posemo	Love, good, happy, :), great	.14***	[0.10, 0.18]	.17***	[0.13, 0.20]
Sad	Miss, lost, sad, sorry, alone	.13***	[0.09, 0.16]	.05*	[0.01, 0.09]
Focus present	Is, have, be, are, get	.12***	[0.09, 0.16]	.04*	[0.00, 0.08]
Adverb	So, just, now, when, about	.11***	[0.08, 0.15]	.04*	[0.01, 0.08]
Focus future	Will, going, then, hope, gonna	.11***	[0.07, 0.15]	.06**	[0.02, 0.09]
Cogproc	All, not, but, if, or	.11***	[0.07, 0.14]	.03	[0.00, 0.07]
Posemo	Love, good, happy, :), great	.14***	[0.10, 0.18]	.17***	[0.13, 0.20]
Affect	Love, good, happy, :), great	.17***	[0.13, 0.20]	.15***	[0.12, 0.19]
Ppron	I, my, you, me, your	.18***	[0.14, 0.21]	.15***	[0.11, 0.19]
Social	You, your, we, love, they	.10***	[0.07, 0.14]	.15***	[0.11, 0.18]
Family	Family, baby, mom, dad, son	.10***	[0.07, 0.14]	.13***	[0.09, 0.16]
Pronoun	I, my, you, it, this	.15***	[0.12, 0.19]	.13***	[0.09, 0.16]
Affiliation	We, love, our, friends, family	.09***	[0.05, 0.12]	.12***	[0.09, 0.16]
Certain	All, never, ever, always, every	.10***	[0.06, 0.13]	.12***	[0.09, 0.16]
Religious	God, hell, pray, holy, praying	.04*	[0.00, 0.08]	.11***	[0.08, 0.15]
I	I, my, me, i'm im	.17***	[0.13, 0.20]	.10***	[0.06, 0.13]

Note. All results Benjamini Hochberg *p* corrected (*** $p < .001$. ** $p < .01$. * $p < .05$). CI = confidence interval; LIWC = Linguistic Inquiry Word Count.

Top 10 Positively Correlated LIWC Categories With Empathy (First 10 Lines) and Compassion (Last 10 Lines)

Next, we considered the same feature set while controlling for the other construct. Several linguistic features were associated with compassion without empathy and empathy without compassion using LIWC (Table 3). Compassion controlling for empathy is more strongly associated with positive emotion words and negatively associated with negative emotion words (β = −0.05, $p < .01$; see Table S4 in the online supplemental materials for correlations with the negative emotions LIWC category), whereas empathy controlling for compassion is associated with both positive and negative emotion words (β = 0.07, $p < .001$). Compassion controlling for empathy is associated with words in the social category, involving other-oriented words such as "you, your, we, they, love," while empathy controlling for compassion is not. This may suggest a greater other-focus on compassion controlling empathy, supporting our theoretical framework. This increased other-focus in the language is further supported by more frequent mentions of the family (family, baby, mom, dad) and social affiliations in general (we, love, our, family). Furthermore, those higher in compassion controlling for empathy more often mentioned positive affect (love, good, happy) and religion words. Empathy controlling for compassion, on the other hand, is associated with expressing words in the first-person pronoun category, which is often indicative of a greater self-focus. The next highest correlation with empathy controlling for compassion is the present tense category followed by sadness. Together, these linguistic findings may suggest that those higher in empathy controlling for compassion are focused more on themselves in the here and now, whereas those higher in compassion controlling for empathy are more focused on social relations and positive emotions.

Table 3

Top 10 Positively Correlated LIWC Categories With Compassion Controlling for Empathy (First 10 Lines) and Empathy Controlling for Compassion (Last 10 Lines)

		Compassion without empathy		Empathy without compassion	
Category	Words	β	95% CI	β	95% CI
Posemo	Love, good, happy, :), great	.10***	[0.06, 0.14]	.05**	[0.01, 0.09]
Social	You, your, we, love, they	.10***	[0.06, 0.13]	.03	[−0.01, 0.06]
Religious	God, hell, pray, holy, praying	.09***	[0.06, 0.13]	−.02	[−0.06, 0.02]
Affiliation	We, love, our, friends, family	.08***	[0.04, 0.12]	.02	[−0.02, 0.06]
Certain	All, never, ever, always, every	.08***	[0.04, 0.11]	.03	[−0.01, 0.07]
Family	Family, baby, mom, dad, son	.08***	[0.04, 0.11]	.04	[0.00, 0.07]
Drives	We, up, get, love, good	.07***	[0.03, 0.11]	−.01	[−0.05, 0.03]
Affect	Love, good, happy, :), great	.07***	[0.03, 0.11]	.09***	[0.05, 0.12]
She/he	He, her, she, his, him	.07***	[0.03, 0.11]	−.02	[−0.05, 0.02]
Male	He, his, him, man, boy	.07***	[0.03, 0.10]	−.02	[−0.05, 0.02]
I	I, my, me, i'm im	.01	[−0.03, 0.05]	.12***	[0.08, 0.15]
Focus present	Is, have, be, are, get	−.03	[−0.06, 0.01]	.10***	[0.07, 0.14]
Ppron	I, my, you, me, your	.06**	[0.02, 0.10]	.10***	[0.06, 0.14]
Sad	Miss, lost, sad, sorry, alone	−.02	[−0.05, 0.02]	.10***	[0.06, 0.14]
Discrep	If, want, need, would, could	−.04	[−0.07, 0.00]	.09***	[0.05, 0.13]
Verb	Is, have, be, are, was	−.03	[−0.06, 0.01]	.09***	[0.05, 0.13]
Adverb	So, just, now, when, about	−.02	[−0.05, 0.02]	.09***	[0.05, 0.13]
Cogproc	All, not, but, if, or	−.02	[−0.06, 0.02]	.09***	[0.05, 0.13]
Pronoun	I, my, you, it, this	.05**	[0.01, 0.08]	.09***	[0.05, 0.13]
Affect	Love, good, happy, :), great	.07***	[0.03, 0.11]	.09***	[0.05, 0.12]

Note. Small effect sizes are typical for LIWC associations (see Pang et al., 2019, Park et al., 2015). CI = confidence interval. Correlations are controlled for the other construct (e.g., empathy without compassion is empathy controlled for compassion) and Benjamini Hochberg *p* corrected (*** $p < .001$. ** $p < .01$. * $p < .05$).

Top 10 Positively Correlated LIWC Categories With Compassion Controlling for Empathy (First 10 Lines) and Empathy Controlling for Compassion (Last 10 Lines)

DLA shows results consistent with LIWC (Figures 2 and 3). Empathy controlling for compassion is associated with words in line with self-focus (me, my, myself), and topics related to negative affect (bleh, (hate), negative states (miss, missing), physical problems (hurts, bad, body), emotional pain and numbness (feeling, tired, numb), and bad mood (mood, bad, crappy, ruin). Compassion controlling for empathy, on the other hand, is associated with expressions of prayers and religiosity (god, blessing, prayer), expressing gratitude (thank you, blessed), and close relationships (family), with topics supporting this impression revolving around love, caring, and giving (loved, giving, caring), resilience and successfully overcoming obstacles (strength, overcome, challenges, trials, faith), love and affection (love, friend, sister), positive affect (fantastic, great, day), and positive perceptions of the world or generalized optimism (world, wonderful, perfect). While the topics in Figure 3 are manually selected, a larger sample of DLA topic results (without manual selection) is shown in Table S3 in the online supplemental materials.

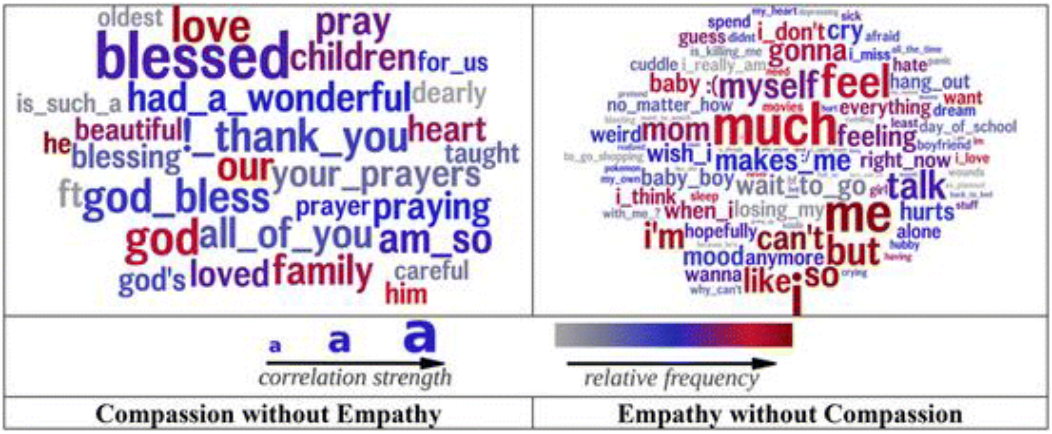

Figure 2. Note. Size of the word indicates the strength of the correlation (Pearson r; Compassion r range: .06–.10; Empathy r range: .05–.11); color indicates the relative frequency of usage. Underscores (_) connect

words of multiword phrases. All correlations are controlled for the other construct (e.g., empathy without compassion is empathy controlled for compassion) and are significant at p < .05 after correcting for multiple comparisons.

Figure 3. Note. Topics represented as the 15 most prevalent words; word size and color indicate word prevalence. All correlations are controlled for the other construct (i.e., empathy without compassion is empathy controlled for compassion) and significant at p < 0.05 after correcting for multiple comparisons.

We then explored how *language* related to empathy (controlling for compassion) and compassion (controlling for empathy) is associated with personality by correlating related language topics with personality factors (as measured by surveys). Figure 4 shows the association of topics with the five factors of personality. In general, people who frequently use empathy controlling for compassion topics tend to be more neurotic, less conscientious, and less agreeable, which would be in line with more commonly experiencing stress and emotional turmoil. Topics high in compassion controlling for empathy are the opposite: they are used more by people higher in conscientiousness, emotional stability, extraversion, and agreeableness.

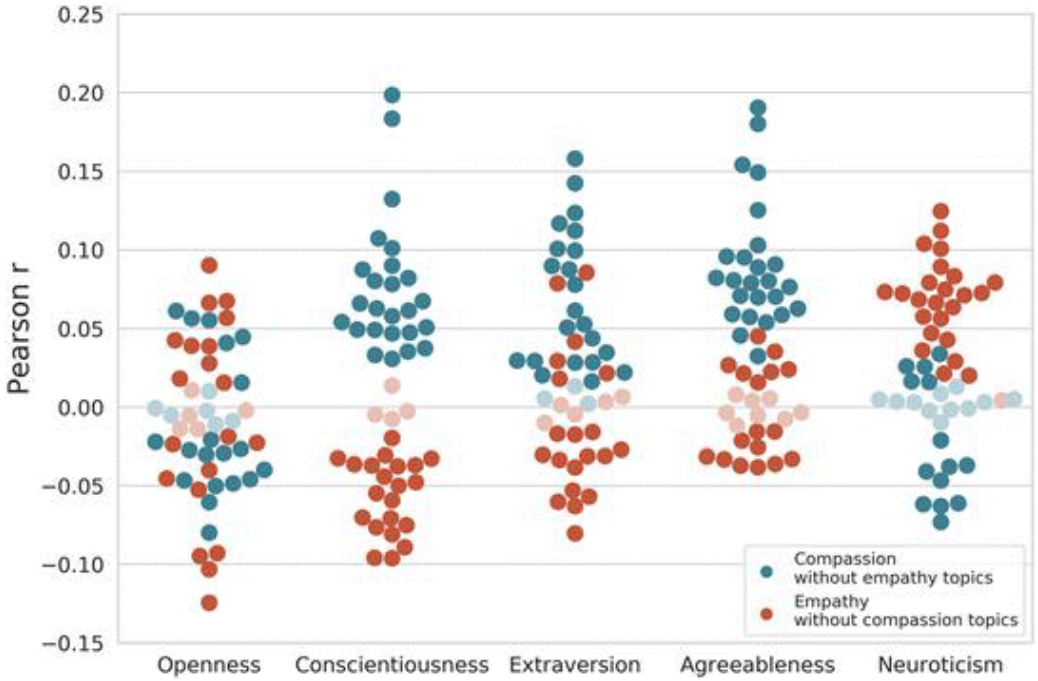

Figure 4. Note. Lighter dots represent nonsignificant correlations after correcting for multiple comparisons (p > .05).

Table 4 shows the sociodemographic and health correlates of empathy controlling for compassion and compassion controlling for empathy. Empathy controlling for compassion is positively correlated with being female and negatively correlated with age, income, and education. Empathy controlling for compassion is associated with more stress, more mental health problems, more sick days, more drinking, and fewer donations to charity. Compassion controlling for empathy is associated with higher age and education, lower stress, better health, more exercise, less drinking, and more charitable donations.

Table 4

Correlates With Demographics (Gender Is Encoded Female=1, Male=0), Socioeconomics, Health, and Prosocial Behaviors

Categories	*N*	*M*	*SD*	Min	Max	Skew	Compassion without empathy β [95% CI]	Empathy without compassion β [95% CI]
Demographics								
Age	2,931	39.04	12.72	18	85.00	0.68	.12*** [0.08, 0.16]	−.24*** [−0.27, −0.21]
Gender	2,931	0.59	0.49	0	1	−0.35	.27*** [0.18, 0.35]	.18*** [0.09, 0.27]
Socioeconomics								
Income	2,931	2.96	1.09	1	6.00	0.55	−.01 [−0.04, 0.03]	−.03 [−0.06, 0.01]
Education	2,931	3.84	1.75	1	8.00	0.36	−.07** [−0.11, −0.03]	−.03 [−0.07, 0.01]
Health								
Perceived stress	2,931	2.39	1.16	1	5.00	0.48	−.14*** [−0.18, −0.10]	.36*** [0.33, 0.39]
Mental health Problems	2,597	1.25	1.29	0	3.83	0.38	.03 [−0.01, 0.07]	.19*** [0.14, 0.23]
Good physical health	2,570	0.92	1.20	0	5.90	0.96	.08** [0.04, 0.13]	.00 [−0.04, 0.05]
Sick days	2,931	2.80	0.75	1	5.00	0.15	−.02 [−0.05, 0.02]	.12*** [0.08, 0.15]
Overweight	2,931	1.05	0.38	0	1.61	−0.72	−.04 [−0.08, −0.00]	.01 [−0.03, 0.05]
Lifestyle								
Exercise	2,931	3.05	1.05	1	5.00	0.17	.06* [0.02, 0.09]	.04 [0.00, 0.07]
Drinking on weekdays	2,931	3.76	2.04	1	8.00	0.57	−.08*** [−0.11, −0.04]	.08*** [0.04, 0.11]
Prosocial behaviors								
Volunteering	2,931	0.39	0.55	0	2.08	1.21	.04 [0.00, 0.07]	.03 [−0.01, 0.07]
Donate to charity	2,931	0.74	0.53	0	1.61	−0.25	.12*** [0.09, 0.16]	−.05* [−0.09, −0.01]

Note. Reported standardized βs with 95% confidence interval in square bracket. CI = confidence interval. Correlations are controlled for the other construct (i.e., empathy or compassion) and Benjamini Hochberg *p* corrected (*** $p < .001$. ** $p < .01$. * $p < .05$).

Correlates With Demographics (Gender Is Encoded Female=1, Male=0), Socioeconomics, Health, and Prosocial Behaviors

General Discussion

While compassion and empathy correlate highly with one another (r = .53, p < .001, CI = [0.50, 0.56])—they nonetheless show a different pattern of results across both language and life outcomes when controlling for the other construct. When looking at correlates of high empathy controlling for compassion, we find an association with poor physical and mental health, missing workdays due to illness, and with greater perceived stress. We also find a small negative association with charitable donations. For compassion controlling for empathy, the reverse is true—we find better physical and mental health and habits, as well as a positive correlation with charitable giving. Hence, while compassion and empathy correlate highly with one another, they do indeed have unique predictive validity once each construct is isolated (i.e., compassion without empathy and empathy without compassion).

Linguistic topics related to compassion (without empathy) and empathy (without compassion) show clear relationships with four of the five personality factors. Topics related to compassion without empathy are marked by higher conscientiousness, extraversion, agreeableness, and emotional stability. Empathy without compassion topics are more associated with introversion and are also moderately associated with neuroticism and lower conscientiousness. The association of low emotional stability and conscientiousness is also in line with prior research that found “distress,” a construct with important parallels to empathy, being associated with fleeing from a helping situation (Batson et al., 1987) and with lower helping (Jordan et al., 2016; Schroeder et al., 1988; Twenge et al., 2007; and others).

In sum, it appears that compassion without empathy and empathy without compassion are at least somewhat distinct and have unique predictive validity in personality, health, and prosocial behavior. While the mechanisms through which these different relationships occur remain unknown, some previous work bears on this issue. As mentioned, other work has found that merely focusing on others resulted in more intentions to help others (Bloom, 2017; Davis, 1983; Jordan et al., 2016), which helps to explain the relationship between the more other-focused compassion and donation behavior that we observed.

Focus on others may involve other psychological processes that can help to explain the findings. Positive affect is associated with other focus, and vice versa (Batson et al., 1987; Buffone et al., 2017; Singer & Klimecki, 2014). Relatedly, positive affect has been associated with adaptive interpersonal approach behavior (Singer & Klimecki, 2014). Decades of research on positive emotion has demonstrated that positive emotion can induce health-related and prosocial behaviors (Fredrickson, 1998, 2000, 2001, 2013). But rather than positive emotion boosting the prosocial and health-related relationship with compassion, it may be more the case that the negative emotion associated with empathy has a suppressing effect on health-related and prosocial behavior. It has been found that negative emotions can be induced by feeling others' pain (Andreychik & Migliaccio, 2015). Some have postulated a motivating effect to get rid of vicarious negative emotion through acting on the behalf of individuals who are suffering (Cialdini, 1991; also see Batson, 2017).

This study was limited in several ways. First, it relies on self-report data. While other forms of data would be ideal, empathy and compassion research is currently primarily reliant on self-report methods. Future research should endeavor to operationalize these constructs through other means. Second, this sample was collected online. Research has shown that online samples are generally reliable and representative (Buhrmester et al., 2016), but these findings should be examined in other, more diverse samples. Third, the linguistic analysis relies on both social media use in general and a minimum number of words posted on social media. Furthermore, there may be differences in empathy, compassion, and personality across populations who use social media frequently versus those who do not. Thus, these results may not generalize to populations who do not use social media.

It is important to note that like with all language-derived personality measures, there are potential ethical issues in the use of empathic personality for decisions involving pricing or denying health insurance. Fortunately, the impact of empathic personality on health is likely to be too small to be used to meaningfully predict someone's future health status.

When breaking empathy and compassion apart and controlling for the other construct, language provides interesting insights into the realities of those who are high in compassion or empathy. Those high in empathy without compassion are more self-focused and experience a greater mixture of positive and negative affect than those high in compassion. Those high in compassion without empathy are other-focused and discuss topics around strength and resilience. Empathy without compassion is related to worse health and is negatively related to charitable donations while compassion without empathy corresponds to both better health and more charitable donations. Such findings support a more critical perspective with regard to the moral power of empathy and a commensurably greater emphasis on the role of compassion.

Footnotes

[1] This survey was part of a larger survey including the Sense of Control scale (Lachman and Weaver, 1998).

References

Andreychik, M. R., & Migliaccio, N. (2015). Empathizing with others' pain versus empathizing with others' joy: Examining the separability of positive and negative empathy and their relation to different types of social

behaviors and social emotions. *Basic and Applied Social Psychology*, *37*(5), 274–291. 10.1080/01973533.2015.1071256

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*, 163–175. 10.1023/B:JADD.0000022607.19833.00444

Batson, C. D. (2010). Empathy-induced altruistic motivation. In M. M. P. R.Shaver (Ed.), *Prosocial motives, emotions, and behavior: The better angels of our nature* (pp. 15–34). American Psychological Association.

Batson, C. D. (2011). *Altruism in humans*. Oxford University Press.

Batson, C. D. (2017). The empathy-altruism hypothesis: What and so what. In E. M.Seppälä, E.Simon-Thomas, S. L.Brown, M. C.Worline, C.Daryl Cameron, & J. R.Doty (Eds.), *The Oxford handbook of compassion science* (pp. 27–40). Oxford University Press.

Batson, C. D., Fultz, J., & Schoenrade, P. A. (1987). Distress and empathy: Two qualitatively distinct vicarious emotions with different motivational consequences. *Journal of Personality*, *55*(1), 19–39. 10.1111/j.1467-6494.1987.tb00426.x

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. 10.1111/j.2517-6161.1995.tb02031.x

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Bloom, P. (2017). *Against empathy: The case for rational compassion*. Random House.

Buffone, A. E., Poulin, M., DeLury, S., Ministero, L., Morrisson, C., & Scalco, M. (2017). Don't walk in her shoes! Different forms of perspective taking affect stress physiology. *Journal of Experimental Social Psychology*, *72*, 161–168. 10.1016/j.jesp.2017.04.001

Buhrmester, M., Kwang, T., & Gosling, S. D. (2016). Amazon's mechanical Turk: A new source of inexpensive, yet high-quality data?*Perspectives on Psychological Science*, *6*(1), 3–5. 10.1177/1745691610393980

Cialdini, R. B. (1991). Altruism or egoism? That is (still) the question. *Psychological Inquiry*, *2*(2), 124–126. 10.1207/s15327965pli0202_3

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, *24*(4), 385–396. 10.2307/2136404

Costa, P. T., Jr., & McCrae, R. R. (2008). The Revised NEO Personality Inventory (NEO-PI-R). In G. J.Boyle, G.Matthews, & D. H.Saklofske (Eds.), The SAGE handbook of personality theory and assessment, Vol. *2*. *Personality measurement and testing* (pp. 179–198). Sage Publications, Inc. https://sk.sagepub.com/reference/hdbk_personalitytheory2/n9.xml

Curtis, B., Giorgi, S., Buffone, A. E., Ungar, L. H., Ashford, R. D., Hemmons, J.,Summers, D., Hamilton, C. & Schwartz, H. A. (2018). Can Twitter be used to predict county excessive alcohol consumption rates?*PloS One*, *13*(4), Article e0194290. 10.1371/journal.pone.0194290

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113–126. 10.1037/0022-3514.44.1.113

Decety, J., & Cowell, J. M. (2014). The complex relation between morality and empathy. *Trends in Cognitive Sciences*, *18*(7), 337–339. 10.1016/j.tics.2014.04.008

Decety, J., & Jackson, P. L. (2006). A social-neuroscience perspective on empathy. *Current Directions in Psychological Science*, *15*(2), 54–58. 10.1111/j.0963-7214.2006.00406.x

De Waal, F. (2010). *The age of empathy: Nature's lessons for a kinder society*. Random House LLC.

Eichstaedt, J. C., Kern, M. L., Yaden, D. B., Schwartz, H. A., Giorgi, S., Park, G., Hagan, C. A., Tobolsky, V. A., Smith, L. K., Buffone, A., Iwry, J., Seligman, M. E. P., & Ungar, L. H. (2021). Closed-and open-vocabulary approaches to text analysis: A review, quantitative comparison, and recommendations. *Psychological Methods*, *26*(4), 398–427. 10.1037/met0000349

Eichstaedt, J. C., Schwartz, H. A., Kern, M. L., Park, G., Labarthe, D. R., Merchant, R. M., Jha, S., Agrawal, M., Dziurzynski, L. A., Sap, M., & Weeg, C. (2015). Psychological language on Twitter predicts county-level heart disease mortality. *Psychological Science*, *26*(2), 159–169. 10.1177/0956797614557867

Eichstaedt, J. C., Smith, R. J., Merchant, R. M., Ungar, L. H., Crutchley, P., Preoţiuc-Pietro, D., Asch, D. A., & Schwartz, H. A. (2018). Facebook Language predicts depression in medical records. *Proceedings of the National Academy of Sciences*, *115*(44), 11203–11208. 10.1073/pnas.1802331115

Fredrickson, B. L. (1998). Cultivated emotions: Parental socialization of positive emotions and self-conscious emotions. *Psychological Inquiry*, *9*(4), 279–281. 10.1207/s15327965pli0904_4

Fredrickson, B. L. (2000). Cultivating positive emotions to optimize health and well-being. *Prevention & Treatment*, *3*(1), Article 1a.

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American Psychologist*, *56*(3), Article 218. 10.1037/0003-066X.56.3.218

Fredrickson, B. L. (2013). Positive emotions broaden and build. In P.Devine & A.Plant (Eds.), *Advances in experimental social psychology* (Vol. *47*, pp. 1–53). Academic Press.

Gelfand, A. E., & Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, *85*(410), 398–409. 10.1080/01621459.1990.10476213

Hoffman, M. L. (2008). Empathy and prosocial behavior. In M.Lewis, J.Haviland-Jones, & L. F.Barrett (Eds.), *Handbook of emotions* (Vol. *3*, pp. 440–455). The Guilford Press.

Hojat, M., Gonnella, J. S., Nasca, T. J., Mangione, S., Vergare, M., Magee, M. (2002). Physician empathy: Definition, components, measurement, and relationship to gender and specialty. *American Journal of Psychiatry*, *159*(9), 1563–1569. 10.1176/appi.ajp.159.9.1563

Hume, D. (1978). *A treatise of human nature*. John Noon. (Original work published in 1739).

Jordan, M. R., Amir, D., & Bloom, P. (2016). Are empathy and concern psychologically distinct?*Emotion*, *16*(8), 1107–1116. 10.1037/emo0000228

Kern, M. L., Eichstaedt, J. C., Schwartz, H. A., Park, G., Ungar, L. H., Stillwell, D. J., Kosinski, M., Dziurzynski, L., & Seligman, M. E. (2014). From “sooo excited!!!” to “so proud”: Using language to study development. *Developmental Psychology*, *50*(1), 178–188. 10.1037/a0035048

Kern, M. L., Park, G., Eichstaedt, J. C., Schwartz, H. A., Sap, M., Smith, L. K., & Ungar, L. H. (2016). Gaining insights from social media language: Methodologies and challenges. *Psychological Methods*, *21*(4), 507–525. 10.1037/met0000091

Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., & Stillwell, D. (2015). Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American Psychologist*, *70*(6), 543–556. 10.1037/a0039210

Lachman, M. E., & Weaver, S. L. (1998). The sense of control as a moderator of social class differences in health and well-being. *Journal of Personality and Social Psychology*, *74*(3), 763–773. 10.1037/0022-3514.74.3.763

McCallum, A. K. (2002). *Mallet: A machine learning for language toolkit*. http://mallet.cs.umass.edu

Morelli, S. A., Lieberman, M. D., & Zaki, J. (2015). The emerging study of positive empathy. *Social and Personality Psychology Compass*, *9*(2), 57–68. 10.1111/spc3.12157

Pang, D., Eichstaedt, J. C., Buffone, A., Slaff, B., Ruch, W., & Ungar, L. H. (2019). The language of character strengths: Predicting morally valued traits on social media. *Journal of Personality*, *88*(2), 287–306. 10.1111/jopy.12491

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., Ungar, L. H., & Seligman, M. E. (2015). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology*, *108*(6), 934–952. 10.1037/pspp0000020

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*.

Preoţiuc-Pietro, D., Eichstaedt, J., Park, G., Sap, M., Smith, L., Tobolsky, V., Andrew Schwartz, H. & Ungar, L. (2015). The role of personality, age, and gender in tweeting about mental illness. In *Proceedings of the 2nd workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality* (pp. 21–30).

Prinz, J. (2011). Is empathy necessary for morality. In P.Goldie & A.Coplan (Eds.), *Empathy: Philosophical and psychological perspectives* (Vol. *1*, pp. 211–229). Oxford University Press.

Schroeder, D. A., Dovidio, J. F., Sibicky, M. E., Matthews, L. L., & Allen, J. L. (1988). Empathic concern and helping behavior: Egoism or altruism?*Journal of Experimental Social Psychology*, *24*(4), 333–353. 10.1016/0022-1031(88)90024-8

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Lucas, R. E., Agrawal, M., Park, G., Lakshmikanth, S., Jha, S., Seligman, M. E. P., & Ungar, L. (2013a). Characterizing geographic variation in well-being using tweets. In *Seventh international AAAI conference on weblogs and social media*.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M. E. P., & Ungar, L. H. (2013b). Personality, gender, and age in the

language of social media: The open-vocabulary approach. *PloS One*, *8*(9), Article e73791. 10.1371/journal.pone.0073791

Schwartz, H. A., Giorgi, S., Sap, M., Crutchley, P., Ungar, L., & Eichstaedt, J. (2017). Dlatk: Differential language analysis toolkit. In *Proceedings of the 2017 conference on empirical methods in natural language processing: system demonstrations* (pp. 55–60).

Schwartz, H. A., Sap, M., Kern, M. L., Eichstaedt, J. C., Kapelner, A., Agrawal, M., Blanco, E., Dziurzynski, L., Park, G., Stillwell, D. & Kosinski, M. (2016). Dlatk: Differential language analysis toolkit Predicting individual well-being through the language of social media. In *Biocomputing 2016: Proceedings of the pacific symposium* (pp. 516–527).

Singer, T., & Klimecki, O. M. (2014). Empathy and compassion. *Current Biology*, *24*, R875–R878. 10.1016/j.cub.2014.06.054

Smith, A. (1976). *The theory of moral sentiments*. Liberty Classics. (Original work published in 1759).

Toi, M., & Batson, C. D. (1982). More evidence that empathy is a source of altruistic motivation. *Journal of Personality and Social Psychology*, *43*(2), 281–292. 10.1037/0022-3514.43.2.281

Twenge, J. M., Baumeister, R. F., DeWall, C. N., Ciarocco, N. J., & Bartels, J. M. (2007). Social exclusion decreases prosocial behavior. *Journal of Personality and Social Psychology*, *92*(1), Article 56. 10.1037/0022-3514.92.1.56

Yaden, D. B., Eichstaedt, J. C., Kern, M. L., Smith, L. K., Buffone, A., Stillwell, D. J., Kosinski, Michal, Ungar, Lyle H., Seligman, Martin EP & Schwartz, H. A. (2018). The language of religious affiliation: Social, emotional, and cognitive differences. *Social Psychological and Personality Science*, *9*(4), 444–452. 10.1177/1948550617711228

Yaden, D. B., Eichstaedt, J. C., Schwartz, H. A., Kern, M. L., Le Nguyen, K., Wintering, N. A., Hood, R. W., Newberg, A. B. (2016). The language of ineffability: Linguistic analysis of mystical experiences. *Journal of Religion and Spirituality*, *8*(3), 244–252. 10.1037/rel0000043

Yaden, D. B., Giorgi, S., Kern, M. L., Adler, A., Ungar, L. H., Seligman, M. E., & Eichstaedt, J. C. (2021). Beyond beliefs: Multidimensional aspects of religion and spirituality in language. *Psychology of Religion and Spirituality*. 10.1037/rel0000408

Zaki, J. (2016). Empathy is a moral force. In K. G.Gray (Ed.), *The atlas of moral psychology* (pp. 49–50). Guilford Press.

Submitted: October 1, 2020 *Revised:* November 19, 2022 *Accepted:* November 22, 2022

Source: Emotion. Vol. 24. (1), Feb, 2024 pp. 106-115)
Accession Number: 2023-72671-001
Digital Object Identifier: 10.1037/emo0001205