Supplemental Material

Participants

Recruitment was done across two separate surveys, both using the Qualtrics platform. The first survey was started 41,369 times. A total of 376 participants were less than 18 and therefore not permitted to continue. The number of participants who did not want to share Facebook data or did not meet the minimum Facebook requirements were 25,058 and 11,476, respectively. A further 590 non-Male participants were removed in order to reach a roughly Male/Female gender balance. Thus, a total of 3,869 participants were able to begin the survey. Next, 506 participants failed attention checks and 1,136 did not complete the survey (i.e., a complete survey is one in which the participant at least saw the final question). This gives a total of 2,227 who completed the survey, though this does not imply every question was answered (including empathy and compassion items). Demographic questions were asked at the end of the survey, and therefore we are not able to assess if there are differences between the sample that completed the survey and those who did not.

The second survey was started 24,802 times, out of which 1,673 participants consented and met the Facebook requirements (i.e., at least 500 words across their status updates and at least 5 status updates within the last 180 days). A total of 1,433 passed all attention checks. Again, demographic questions were asked at the end of the survey, thus we are unable to assess demographic differences between participants who completed and did not complete the survey.

We combined the 2,227 and 1,433 responses from each survey above. From this, we further removed 42 entries for non-Male/Female gender (33 non-responses and 9 “other”), 64 entries with no reported age, and 11 entries with reported age less than 18. Another 26 entries were removed for not completing the empathy and compassion items. Finally, 553 entries were removed for having completed the survey more than once or duplicate Facebook accounts were used across multiple survey responses. The final sample consisted of 2,931 participants (used for non-language analysis). This was further reduced to 2,781 participants after cleaning Facebook data (removing duplicate posts and removing non-English posts using the langid Python package; Lui and Baldwin, 2012).

We also note that while the surveys were started 41,369 and 24,802, we have no way of verifying that they were all distinct participants since no data was collected (i.e., Facebook data is not retained if participants do not meet the requirements nor are they allowed to proceed to the survey). Thus, it may not be correct to say, for example, that roughly 36,000 *distinct* participants were excluded from the first survey due to not meeting our Facebook requirements. The Qualtrics platform contains multiple options to stop spammers from attempting a survey, such as IP blocking and the prevention of multiple submissions (both of which were implemented in the above surveys). Despite this, the Qualtrics platform is still susceptible to spam, hence our inclusion of attention checks and deleting any entries which contain duplicate identifiers (e.g., Facebook handles).

Factor Analysis

	Varimax	Promax

Question	Construct	Original Scale	Scale	Factor 1	Factor 2	Factor 1	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	.63	.17	.63	.06
I sometimes find myself feeling the emotions of the people around me, even if I don't try to feel what they're feeling.				.66	.24	.64	.13
If I see someone fidgeting, I'll start feeling anxious too.				.57	-.02	.61	-.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	.73	.19	.72	.06
Other people's sadness or upset is contagious for me.				.75	.19	.75	.06
When I see someone cry, I am very likely to cry with them.				.64	.23	.61	.13
If I hear about someone very similar to me experiencing a tragedy, I automatically experience their sadness and suffering as my own.				.71	.29	.68	.18
When I imagine myself coming to the scene of an accident, I imagine myself feeling quite distraught and in emotional turmoil. I imagine feeling like falling to pieces and confused and frazzled, unsure how I could best provide help.				.57	-.03	.61	-.14
I often have tender, concerned feelings for people less fortunate than me.	Compassion	IRI - Empathic Concern	1-5	.16	.66	-.02	.69
When I see someone being taken advantage of, I feel kind of protective towards them.				.12	.60	-.04	.62
I am often quite touched by things that I see happen.				.27	.58	.13	.58
I would describe myself as a pretty soft-hearted person.				.26	.55	.13	.54
When I see someone being treated unfairly, I sometimes don't feel very much pity for them.				.12	-.46	.26	-.52
Other people's misfortunes do not usually disturb me a great deal.				-.10	-.51	.04	-.53
Sometimes I don't feel very sorry for other people when they are having problems.				.01	-.44	.13	-.48
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	.33	.66	.16	.65

I often try to imagine how another person must feel with regard to what happened to them.	.32	.64	.17	.63
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.	.43	.57	.30	.54

Table S1. *Exploratory Factor Analysis. Italicized questions not used in final scale.*

Question	Construct	Original Scale	Scale	Varimax		Promax		No Rotation	
				Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	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	.53	.41	.47	.27	.67	.06
I sometimes find myself feeling the emotions of the people around me, even if I don't try to feel what they're feeling.				.39	.35	.32	.26	.52	.01
If I see someone fidgeting, I'll start feeling anxious too.				.08	.71	-.21	.83	.54	-.46
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	.24	.64	.01	.67	.61	-.30
Other people's sadness or upset is contagious for me.				.17	.87	-.17	.99	.72	-.52
When I see someone cry, I am very likely to cry with them.				.35	.45	.24	.39	.56	-.08
If I hear about someone very similar to me experiencing a tragedy, I automatically experience their sadness and suffering as my own.				.31	.64	.10	.64	.66	-.26
I often have tender, concerned feelings for people less fortunate than me.	Compassion	IRI - Empathic Concern	1-5	.71	.14	.81	-.14	.62	.38
When I see someone being taken advantage of, I feel kind of protective towards them.				.52	.19	.56	.00	.51	.22
I am often quite touched by things that I see happen.				.63	.24	.66	.01	.62	.26
I would describe myself as a pretty soft-hearted person.				.55	.30	.54	.12	.61	.16
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	.71	.11	.82	-.17	.60	.40
I often try to imagine how another person must feel with regard to what happened to them.				.80	.15	.91	-.17	.69	.43

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.				.72	.31	.74	.06	.74	.27

Table S2. Scale questions with factor loadings in follow up dataset. Varimax, promax, and unrotated solutions.

Question	Construct	Original Scale	Scale	Varimax	
				Factor 1	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	.58	.25
I sometimes find myself feeling the emotions of the people around me, even if I don't try to feel what they're feeling.				.65	.28
If I see someone fidgeting, I'll start feeling anxious too.				.54	.07
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	.74	.22
Other people's sadness or upset is contagious for me.				.76	.22
When I see someone cry, I am very likely to cry with them.				.64	.24
If I hear about someone very similar to me experiencing a tragedy, I automatically experience their sadness and suffering as my own.				.68	.35
I often have tender, concerned feelings for people less fortunate than me.	Compassion	IRI - Empathic Concern	1-5	.14	.61
When I see someone being taken advantage of, I feel kind of protective towards them.				.09	.57
I am often quite touched by things that I see happen.				.25	.56
I would describe myself as a pretty soft-hearted person.				.24	.52
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	.23	.76
I often try to imagine how another person must feel with regard to what happened to them.				.23	.73
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.				.34	.68

Table S3. Scale questions with factor loadings for Varimax rotation in current dataset.

Additional Language Correlates

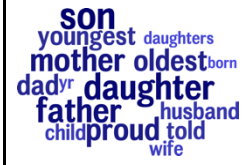
				
.13 [.09, .16]***	.09 [.08, .16]***	.09 [.07, .14]***	.09 [.07, .14]***	.09 [.06, .14]***
				
.09 [.06, .13]***	.09 [.06, .13]***	.09 [.06, .13]***	.08 [.06, .13]***	.08 [.05, .13]***

Figure S3. *Topics most highly distinguishing of compassion without empathy (i.e., compassion controlled for empathy). The size of the word indicates the relative contribution of each word to that topic. Color (darkness) is proportional to size. Reported standardized beta with 95% confidence intervals in square brackets. All correlations significant after correcting for multiple comparisons* (****p < .001, **p < .01, *p < .05*).

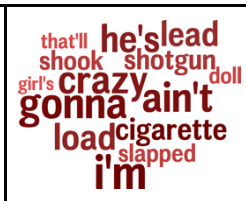
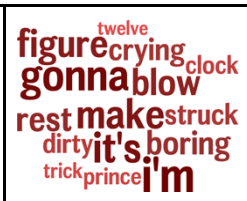
				
.11 [.08, .15]***	.11 [.07, .15]***	.11 [.07, .14]***	.10 [.07, .14]***	.10 [.07, .14]***
				
.10 [.07, .14]***	.10 [.06, .14]***	.10 [.06, .14]***	.10 [.06, .14]***	.10 [.06, .13]***

Figure S4. *Topics most highly distinguishing of empathy without compassion (i.e., empathy controlled for compassion). The size of the word indicates the relative contribution of each word to that topic. Color (darkness) is proportional to size. Reported standardized beta with 95% confidence intervals in square brackets. All correlations significant after correcting for multiple comparisons* (****p < .001, **p < .01, *p < .05*).

	Compassion without empathy		Empathy without compassion	
	β	95% CI	β	95% CI
ACHIEVE	-0.04*	[-0.08, 0.00]	-0.01	[-0.05, 0.02]
ADJ	-0.01	[-0.04, 0.03]	0.05**	[0.01, 0.08]
ADVERB	-0.02	[-0.05, 0.02]	0.09***	[0.05, 0.13]
AFFECT	0.07***	[0.03, 0.11]	0.09***	[0.05, 0.12]
AFFILIATION	0.08***	[0.04, 0.12]	0.02	[-0.02, 0.06]
ANGER	-0.06*	[-0.09, -0.02]	0.01	[-0.03, 0.05]
ANX	0.00	[-0.04, 0.04]	0.07***	[0.04, 0.11]
ARTICLE	-0.06**	[-0.10, -0.02]	-0.08***	[-0.12, -0.05]
ASSENT	0.00	[-0.03, 0.04]	0.03	[-0.01, 0.06]
AUXVERB	-0.01	[-0.05, 0.03]	0.07***	[0.03, 0.11]
BIO	-0.01	[-0.04, 0.03]	0.06**	[0.02, 0.10]
BODY	-0.03	[-0.07, 0.01]	0.04*	[0.01, 0.08]
CAUSE	-0.02	[-0.06, 0.01]	0.03	[0.00, 0.07]
CERTAIN	0.08***	[0.04, 0.11]	0.03	[-0.01, 0.07]
COGPROC	-0.02	[-0.06, 0.02]	0.09***	[0.05, 0.13]
COMPARE	-0.06**	[-0.10, -0.02]	0.04*	[0.00, 0.07]
CONJ	0.02	[-0.02, 0.05]	0.06***	[0.03, 0.1]
DEATH	-0.03	[-0.06, 0.01]	-0.02	[-0.05, 0.02]
DIFFER	-0.05*	[-0.08, -0.01]	0.08***	[0.04, 0.12]
DISCREP	-0.03	[-0.07, 0.00]	0.09***	[0.05, 0.13]
DRIVES	0.07***	[0.03, 0.11]	-0.01	[-0.05, 0.03]
FAMILY	0.07***	[0.04, 0.11]	0.04	[0.00, 0.07]
FEEL	0.00	[-0.04, 0.04]	0.07***	[0.03, 0.10]
FEMALE	0.05**	[0.02, 0.09]	0.02	[-0.01, 0.06]
FILLER	-0.01	[-0.05, 0.03]	0.05**	[0.01, 0.09]
FOCUSFUTURE	0.00	[-0.04, 0.04]	0.08***	[0.04, 0.12]
FOCUSPAST	-0.01	[-0.05, 0.03]	-0.01	[-0.05, 0.03]
FOCUSPRESENT	-0.03	[-0.06, 0.01]	0.10***	[0.07, 0.14]
FRIEND	0.03	[0.00, 0.07]	0.02	[-0.02, 0.06]
FUNCTION	0.00	[-0.04, 0.04]	0.06**	[0.02, 0.10]
HEALTH	0.02	[-0.02, 0.06]	0.05**	[0.01, 0.09]
HEAR	0.02	[-0.02, 0.05]	-0.02	[-0.05, 0.02]
HOME	0.02	[-0.02, 0.05]	0.00	[-0.04, 0.04]
I	0.01	[-0.03, 0.05]	0.12***	[0.08, 0.15]
INFORMAL	-0.01	[-0.05, 0.03]	0.06**	[0.02, 0.10]
INGEST	-0.03	[-0.07, 0.00]	0.00	[-0.03, 0.04]
INSIGHT	-0.01	[-0.05, 0.03]	0.04*	[0.01, 0.08]

INTERROG	-0.02	[-0.05, 0.02]	0.03	[0.00, 0.07]
IPRON	0.00	[-0.04, 0.03]	0.02	[-0.01, 0.06]
LEISURE	-0.05*	[-0.08, -0.01]	0.01	[-0.03, 0.05]
MALE	0.07***	[0.03, 0.10]	-0.01	[-0.05, 0.02]
MONEY	-0.03	[-0.07, 0.01]	-0.05**	[-0.09, -0.01]
MOTION	-0.04*	[-0.08, -0.01]	0.02	[-0.02, 0.06]
NEGATE	-0.03	[-0.06, 0.01]	0.08***	[0.04, 0.11]
NEGEMO	-0.05**	[-0.09, -0.01]	0.07***	[0.03, 0.11]
NETSPEAK	0.02	[-0.02, 0.05]	0.06***	[0.02, 0.10]
NONFLU	-0.02	[-0.06, 0.01]	0.04*	[0.00, 0.08]
NUMBER	-0.03	[-0.07, 0.00]	-0.05**	[-0.09, -0.01]
PERCEPT	0.00	[-0.04, 0.04]	0.03	[-0.01, 0.07]
POSEMO	0.10***	[0.06, 0.14]	0.05**	[0.01, 0.09]
POWER	0.06**	[0.02, 0.09]	-0.07***	[-0.11, -0.04]
PPRON	0.06**	[0.02, 0.10]	0.10***	[0.06, 0.14]
PREP	-0.05*	[-0.08, -0.01]	-0.03	[-0.07, 0.01]
PRONOUN	0.05*	[0.01, 0.08]	0.09***	[0.05, 0.13]
QUANT	0.01	[-0.03, 0.04]	0.02	[-0.02, 0.06]
RELATIV	-0.07***	[-0.10, -0.03]	0.01	[-0.02, 0.05]
RELIG	0.09***	[0.06, 0.13]	-0.02	[-0.06, 0.02]
REWARD	-0.02	[-0.05, 0.02]	0.02	[-0.02, 0.06]
RISK	-0.01	[-0.05, 0.02]	0.00	[-0.04, 0.03]
SAD	-0.02	[-0.05, 0.02]	0.10***	[0.06, 0.14]
SEE	-0.02	[-0.06, 0.01]	0.01	[-0.02, 0.05]
SEXUAL	-0.06**	[-0.09, -0.02]	0.02	[-0.01, 0.06]
SHEHE	0.07***	[0.03, 0.11]	-0.02	[-0.05, 0.02]
SOCIAL	0.09***	[0.06, 0.13]	0.03	[-0.01, 0.06]
SPACE	-0.05**	[-0.09, -0.01]	-0.05**	[-0.08, -0.01]
SWEAR	-0.06**	[-0.09, -0.02]	0.01	[-0.02, 0.05]
TENTAT	-0.06**	[-0.09, -0.02]	0.07***	[0.04, 0.11]
THEY	0.04*	[0.00, 0.08]	-0.07***	[-0.10, -0.03]
TIME	-0.05**	[-0.09, -0.02]	0.05*	[0.01, 0.08]
VERB	-0.03	[-0.06, 0.01]	0.09***	[0.05, 0.13]
WE	0.05**	[0.01, 0.09]	-0.02	[-0.05, 0.02]
WORK	-0.05**	[-0.09, -0.01]	-0.04*	[-0.08, -0.01]
YOU	0.04*	[0.00, 0.08]	0.05**	[0.01, 0.09]

Table S4: *Full LIWC correlations for compassion without empathy and empathy without compassion. All correlations significant after correcting for multiple comparisons* (****p < .001, **p < .01, *p < .05*).

Language Replication

We attempt to replicate the language correlates of empathy and compassion in a second sample. The second sample is a subset (N = 517) of the participants used to replicate the factor analysis (see the section on *Participants and procedures* in the Methods section for a description of this sample). Specifically, the participants used to replicate the factor analysis only needed to have answered the empathy and compassion scale items, whereas the participants in this replication have an additional requirement of at least 500 words across their Facebook post history. Using the thresholds in Eichstaedt et al. (2021), for a sample size of 500, we can expect less than 10 LIWC and ngram correlations and less than 100 topic correlations for psychological constructs similar to personality (as opposed to other outcomes like age and gender).

We attempt to replicate the language correlations using three language categories: (1) LIWC, (2) Facebook topics, (3) ngrams. For all three language categories we repeat the following process. First, we run DLA in the original sample: run a linear regression with empathy (or compassion) as the dependent variable, the individual language feature as an independent variable (e.g., a single LIWC category or single Facebook topic), and include compassion (or empathy) as a covariate. All variables are z-scored (i.e., mean centered with unit standard deviation). We then take the standardized beta's from the independent variable for each regression and create a *M*x1 dimensional vector, where *M* is the number of individual features in the language category (e.g., 2000 Facebook topics). Second, we run DLA in the second sample creating a second *M*x1 dimensional vector of standardized betas. Finally, we run a Pearson correlation between these two *M*x1 vectors. These three steps are repeated for both empathy and compassion. We then report the Pearson r for each language category. This process has previously been used to establish linguistic similarities between constructs in such a way that produces a single, interpretable number when using a large number of language features (Park et al. 2015).

	LIWC	Facebook Topics	1, 2 and 3-grams
Empathy without compassion	0.46	0.52	0.41
Compassion without compassion	0.82	0.77	0.60

Table S5: Only considering features significantly correlated in the original sample.

References

Eichstaedt, J. C., Kern, M. L., Yaden, D. B., Schwartz, H. A., Giorgi, S., Park, G., ... & Ungar, L. H. (2021). Closed-and open-vocabulary approaches to text analysis: A review, quantitative comparison, and recommendations. Psychological Methods, 26(4), 398.

Lui, M., & Baldwin, T. (2012, July). langid. py: An off-the-shelf language identification tool. In *Proceedings of the ACL 2012 system demonstrations* (pp. 25-30).

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., ... &

Seligman, M. E. (2015). Automatic personality assessment through social media language. *Journal of personality and social psychology*, *108*(6), 934.