

## ARTICLE OPEN



# Head versus heart: social media reveals differential language of loneliness from depression

Tingting Liu<sup>1,2</sup>, Lyle H. Ungar<sup>2,3</sup>, Brenda Curtis<sup>1</sup>, Garrick Sherman<sup>2</sup>, Kenna Yadeta<sup>1</sup>, Louis Tay<sup>4</sup>, Johannes C. Eichstaedt<sup>5</sup> and Sharath Chandra Guntuku<sup>3</sup>

We study the language differentially associated with loneliness and depression using 3.4-million Facebook posts from 2986 individuals, and uncover the statistical associations of survey-based depression and loneliness with both dictionary-based (Linguistic Inquiry Word Count 2015) and open-vocabulary linguistic features (words, phrases, and topics). Loneliness and depression were found to have highly overlapping language profiles, including sickness, pain, and negative emotions as (cross-sectional) risk factors, and social relationships and activities as protective factors. Compared to depression, the language associated with loneliness reflects a stronger cognitive focus, including more references to cognitive processes (i.e., differentiation and tentative language, thoughts, and the observation of irregularities), and cognitive activities like reading and writing. As might be expected, less lonely users were more likely to reference social relationships (e.g., friends and family, romantic relationships), and use first-person plural pronouns. Our findings suggest that the mechanisms of loneliness include self-oriented cognitive activities (i.e., reading) and an overattention to the interpretation of information in the environment. These data-driven ecological findings suggest interventions for loneliness that target maladaptive social cognitions (e.g., through reframing the perception of social environments), strengthen social relationships, and treat other affective distress (i.e., depression).

*npj Mental Health Research* (2022)1:16; <https://doi.org/10.1038/s44184-022-00014-7>

## INTRODUCTION

Loneliness, defined as the emotional unpleasant experience when individuals perceive their interpersonal relationships are not up to their expectations<sup>1–3</sup>, is highly prevalent in the general population and people with mental health disorders<sup>3–5</sup>. Loneliness impacts psychological functioning and predicts increased morbidity and mortality<sup>2,6,7</sup>. Multiple prior studies in the psychology, public health, and psychiatry domains report that loneliness is highly correlated with depression (e.g., Pearson's  $r$  of 0.4–0.6<sup>5,8–13</sup>), a common mental disorder characterized by persistent sadness, disturbed sleep and appetite, and anhedonia<sup>14</sup>. Therefore, loneliness is often investigated as a transdiagnostic facilitator together with depression or considered as an adjunct to the treatment of depression<sup>5,8–13</sup> and has been overlooked during clinical encounters<sup>6,15</sup>. To develop interventions specifically targeted at reducing loneliness, we need to distinguish the unique markers of loneliness from depression. The current study uses both psychological assessments and language from Facebook posts to shed light on the distinctive markers of loneliness.

The relationship between loneliness and depression appears to be reciprocal in nature. In studies on causes, consequences, and treatment plans for loneliness, depression is frequently considered to be an important influential factor and a wide-known pathological consequence<sup>16,17</sup>. In studies on depression, loneliness is identified as a risk factor and subsymptom for depression<sup>9,18</sup>. Both loneliness and depression have been linked to social skills deficits, maladaptive social cognition, internalizing distress, and uncontrollable thoughts, including worries and ruminating thoughts<sup>19–21</sup>. In contrast to the well-discussed association between loneliness and depression, limited attention has been given to understanding the unique markers of loneliness

from depression. Initial evidence indicates that loneliness and depression are two related yet separate constructs<sup>22</sup>. For example, a 5-year longitudinal survey study showed that loneliness and depression are statistically separable, and loneliness predicted subsequent changes in depressive symptoms<sup>17</sup>. In addition, other studies suggest that loneliness has its unique symptomology<sup>23</sup>. Even for the mechanisms shared by loneliness and depression, recent findings from self-report surveys suggest that variance emphasis on uncontrolled thoughts and worries might distinguish loneliness from depression<sup>19</sup>. Specifically, loneliness has a unique function in cognition by providing implicit hypervigilance for social threats<sup>24</sup>, although such cognitive bias in the social context was initially evolved to adversely protect individuals from social threats and meaningless social interactions<sup>25</sup>.

It is important to understand such differential features associated with loneliness<sup>26</sup>. Considering the strong association between loneliness and depression, loneliness is often overshadowed in research on depression (e.g., as a subsymptom<sup>18,27</sup>). Such vagueness might lead to the unsuccessful choice of intervention types that target different components in reducing loneliness (e.g., improving social skills, increasing social supports, improving social cognition abilities, etc.)<sup>20</sup>, and might also result in reducing the efficacy of interventions on loneliness. More importantly, loneliness is uniquely linked to negative feelings associated with perceptions of expectations about social relationships<sup>25</sup>, an inherently social need of being human. Being lonely may lead to perceiving social interactions negatively, while being depressed is not necessarily accompanied by such negative expectations.

To understand the mechanisms of loneliness and depression, we utilized individuals' Facebook language in addition to standardized psychometric assessments in the present study.

<sup>1</sup>National Institute on Drug Abuse (NIDA IRP), National Institutes of Health (NIH), Baltimore, MD, USA. <sup>2</sup>Positive Psychology Center, University of Pennsylvania, Philadelphia, PA, USA. <sup>3</sup>Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA, USA. <sup>4</sup>Department of Psychological Sciences, Purdue University, West Lafayette, IN, USA. <sup>5</sup>Department of Psychology, Institute for Human-Centered A.I., Stanford University, Stanford, CA, USA. ✉email: [tingting.liu@nih.gov](mailto:tingting.liu@nih.gov); [sharathg@cis.upenn.edu](mailto:sharathg@cis.upenn.edu)

Language on social media platforms contains rich ecological signals reflecting one's emotions, thoughts, behavioral patterns, and traits<sup>28</sup>, and at the same time, can be quantified and interpreted for subsequent predictions<sup>29</sup>. Social media language has been shown to provide nuanced insights into and reliably predict individuals' psychology and cognition<sup>30,31</sup> as well as revealing the online sequelae of mental health conditions<sup>32–35</sup>. Facebook, as the most widely used social media platform in the United States<sup>36</sup>, provides a unique opportunity to document one's daily life and has been used to extract linguistic features to successfully predict depression<sup>27</sup>, personality<sup>31</sup>, substance use disorder treatment outcomes<sup>37</sup>, and suicide risk<sup>38</sup>.

In the present study, we surveyed a large sample of individuals across the United States using established psychological assessments (UCLA Loneliness 3-item subscale, ULS-3)<sup>39</sup> and clinical standards (Patient Health Questionnaire, PHQ-9)<sup>40</sup>; and found language markers significantly associated with both constructs using open-vocabulary words, phrases, and themes along with the closed-vocabulary psycholinguistic dictionary, Linguistic Inquiry Word Count 2015 (LIWC 2015)<sup>41</sup>. We obtain significant language markers of loneliness after controlling for age and gender, before and after controlling for depression. Our contributions are two-fold. First, we show that loneliness and depression have large overlaps in language markers of protective and risky factors, and we further elucidate unique features of loneliness through language use patterns; second, we demonstrate the predictive utility of language in assessing loneliness and depression. Our findings provide initial evidence from individuals' Facebook language on the manifestation of loneliness and its relationship to depression, which can potentially inform future personalized interventions to reduce loneliness.

## METHODS

### Participants

The current study is the secondary data analysis of the data collected from a larger study, recruited via the Qualtrics Panel. In the original study, United States-based participants received an incentive for answering a series of surveys and were invited to share access to their Facebook status updates. We obtained informed consent to access their Facebook posts and extracted all posts via the Facebook Graph API. Out of 3215 participants recruited, 3043 participants passed the attention check question. Two thousand nine hundred eighty-six participants ( $M(SD)_{age} = 43.24(12)$  year-old, 69.7% female, 63.8% have Bachelor's or higher degrees, see Supplementary Table 1; see segmented demographics and examples of rephrased Facebook posts by depression and loneliness scores in Supplementary Table 2) who completed ULS-3, PHQ-9, and the key demographic questions (e.g., age, gender) to present study and had active accounts with more than 500 words on Facebook, were included in the data analysis. The gender in our analysis was categorized into two categories ("is female" vs. other). The original study received approval from the Institutional Review Board (IRB) of Purdue University, and the secondary data analysis in the present study was exempted by the IRB of the University of Pennsylvania.

### Materials

**3-item UCLA loneliness (ULS-3)<sup>39</sup>.** The 3-item UCLA loneliness scale consists of three questions from the 20-item Revised UCLA Loneliness scale<sup>42</sup> ("How often do you feel that you lack companionship?," "How often do you feel left out?," "How often do you feel isolated from others?") using the 4-point Likert scoring system to capture more variance of loneliness<sup>43,44</sup> (1 = "Never", 2 = "Rarely", 3 = "Sometimes", 4 = "Often"). The total score of ULS-3 ranges from 3–12, with higher scores representing greater loneliness. This scale has been widely used in previous literature

and displayed good reliability<sup>45</sup>. In the current sample, ULS-3 showed good reliability (Cronbach's  $\alpha = 0.81$ ).

**Patient Health Questionnaire-9 (PHQ-9)<sup>40</sup>.** The PHQ-9 is a 9-item questionnaire based on the nine criteria for depression disorders in the DSM-IV (i.e., "little interest or pleasure in doing things;"). It provides the diagnosis of depression and assesses the severity of the symptoms in the past two weeks using a 4-point scale (from "Not at all" to "Nearly every day"). Previous studies have identified its great internal reliability ( $\alpha > 0.85$ ) and test-retest reliability<sup>46,47</sup>. In the current sample, PHQ-9 showed excellent reliability (Cronbach's  $\alpha = 0.90$ ).

### Linguistic attributes

We characterized participants' Facebook posts (3,459,854 Facebook posts) using two sets of language features: (a) dictionary-based psycholinguistic features, and (b) open-vocabulary topics. Following prior studies<sup>27</sup>, we included participants' entire Facebook language to yield interpretable and fine-grained language variables and results.

**Closed vocabulary.** We first utilized an established language dictionary based on categories of words developed by psychologists, Linguistic Inquiry and Word Count (LIWC 2015)<sup>41</sup>. This top-down (theory-driven) approach has been widely used in past language research and proved to be accurate in detecting linguistic patterns associated with psychological traits such as personality<sup>29</sup>. We calculated the relative frequency of each LIWC category by summing up the within-participant word frequencies within each LIWC category.

**Open vocabulary.** We then used an open-vocabulary approach to extract words and phrases (1–3-grams) and used Latent Dirichlet Allocation topic modeling (LDA)<sup>48</sup> to generate data-driven linguistic features called topics. We split ("tokenized") the Facebook posts into words, punctuation, and emoticons. Facebook posts were tokenized using *happierfuntokenizing* (DLATK/*happierfuntokenizing*, 2017), which improves over most tokenizers for tokenizing emoticons. All words used by less than 1% of users were removed from the analysis to remove uncommonly used words (outliers). This resulted in 4143 unique 1–3-grams.

The LDA generative model assumes that posts are generated by a combination of topics and that topics are a distribution of words. Since the words in a post are known, topics, which are latent variables, can be estimated through Gibbs sampling. An example of such a model is the following set of words (e.g., 'tuesday', 'monday', and 'wednesday'), which cluster together days of the week by discovering their similar distributional properties across posts. We used an open-source set of 2000 topics trained on a corpus of over 20 million Facebook statuses<sup>31</sup>. We calculated the topic distribution of each user aggregated across all posts.

**BERT embeddings.** We also used Bidirectional Encoder Representations from Transformers (BERT), a pre-trained contextual word embedding model, to generate numeric vector representations of users' language. Because the BERT embedding of each word differs depending on what other words are used near it, we expect these models to capture contextual semantic information in our Facebook language data that is lost with other methods. To calculate our user vectors, we represent each word by its 10th layer in the BERT model; these vectors are averaged to produce a message embedding, and messages are finally averaged to produce user-level vector representations.

### Identifying differentially expressed language features for loneliness

We designed this as a person-level analysis and used each language feature dimension (LIWC, words/phrases, and LDA topics) as an independent variable in an OLS linear regression

model to predict individuals' loneliness with age and gender as controls, and then with depression as an additional control. By conventional linguistic analysis, we used a  $p$ -value of  $<0.01$  for LIWC and LDA topics and  $<0.05$  for words and phrases. All  $p$ -values were corrected using the Benjamini–Hochberg False Discovery Rate correction (BH-FDR correction).

### Building prediction models for loneliness using language

We evaluated how well Facebook language could predict loneliness and depression. We treated each linguistic feature set (LIWC 2015 categories, 1–3-grams, LDA topics, and BERT embeddings) as independent variables, and treated self-report loneliness and depression scores as continuous dependent variables in predictive models. Each feature set was considered independently to enable a comparative analysis of their effectiveness. The predictive model was trained using linear regression with ridge regularization, on the training set and evaluated on a test set to avoid overfitting, via cross-validation. Hyperparameter selection was performed within the cross-validation. We report Pearson's  $r$  on an out-of-sample 10-fold cross-validation setting to showcase the predictive power.

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## RESULTS

### Interoutcome survey correlations

We conducted Pearson's correlations among age, loneliness, and depression. We found that consistent with prior work, loneliness is significantly increased with higher levels of depression but decreased with people's age<sup>5</sup>. To examine the shared and unique associations between them, we conducted a series of cross-control partial correlations<sup>49</sup> (Table 1). When controlling for loneliness, the correlations between age and depression remained statistically unchanged ( $p < 0.001$ ). When controlling for depression, the correlation between loneliness and age became insignificant ( $p > 0.1$ ). Results of Welch  $t$ -tests comparing the gender differences for loneliness and depression showed no effects of gender on both (both  $|t| < 1.7$ ,  $p > 0.1$ ).

### Language correlates with loneliness and depression

We correlated loneliness and depression with linguistic features extracted from participants' Facebook posts via two approaches: (a) dictionary-based closed-vocabulary (LIWC 2015)<sup>41</sup>, and (b) open-vocabulary-based words and phrases (1–3-grams) and Latent Dirichlet Allocation (LDA) Facebook topics<sup>48</sup>. We controlled for age and gender in all analyses and then added depression as an additional control.

### Loneliness and depression have large overlaps in language markers

Across all approaches, we found that loneliness has a large amount of shared risk and protective language markers with depression. That is, for the linguistic markers that were significantly correlated with loneliness, 87.1% LIWC categories ( $p < 0.01$ ), 51.8% words and phrases ( $p < 0.05$ ), 74.9% LDA topics ( $p < 0.01$ ), with Benjamini–Hochberg False Discovery Rate (BH-FDR) correction, were also significantly correlated with depression. For example, among the 31 LIWC categories that were significantly correlated with loneliness ( $r_{\text{positive}} = [0.054, 0.122]$ ,  $r_{\text{negative}} = [-0.055, -0.122]$ ), 27 were also significantly correlated with depression at the same significance level ( $r_{\text{positive}} = [0.054, 0.164]$ ,  $r_{\text{negative}} = [-0.073, -0.094]$ ,  $p < 0.01$  with BH-FDR correction, see the full list of LIWC categories in Supplementary Table 3).

**Table 1.** Pearson's and partial correlations between loneliness, depression, and age.

Controlled variable	Loneliness $r$ [95% CI]	Depression $r$ [95% CI]	Age $r$ [95% CI]
None			
Loneliness	1		
Depression	0.554 [0.53, 0.58]***	1	
Age	-0.139 [-0.17, -0.10]***	-0.256 [-0.29, -0.22]***	1
Age			
Loneliness	1		
Depression	0.541 [0.52, 0.57]***	1	
Depression			
Loneliness	1		
Age	0.004 [-0.03, 0.04]		1
Loneliness			
Depression		1	
Age		-0.218 [-0.25, -0.18]***	1

Numbers in the tables are correlation coefficients with the 95% confidence interval in square brackets: [lower bound, upper bound]. Loneliness was assessed using 3-item UCLA Loneliness Scale; Depression was assessed using Patient Health Questionnaire-9. All correlations were bivariate correlations. All  $p$ -values were corrected using the Benjamini–Hochberg correction.

95% CI 95% confidence interval, L lower, U upper.

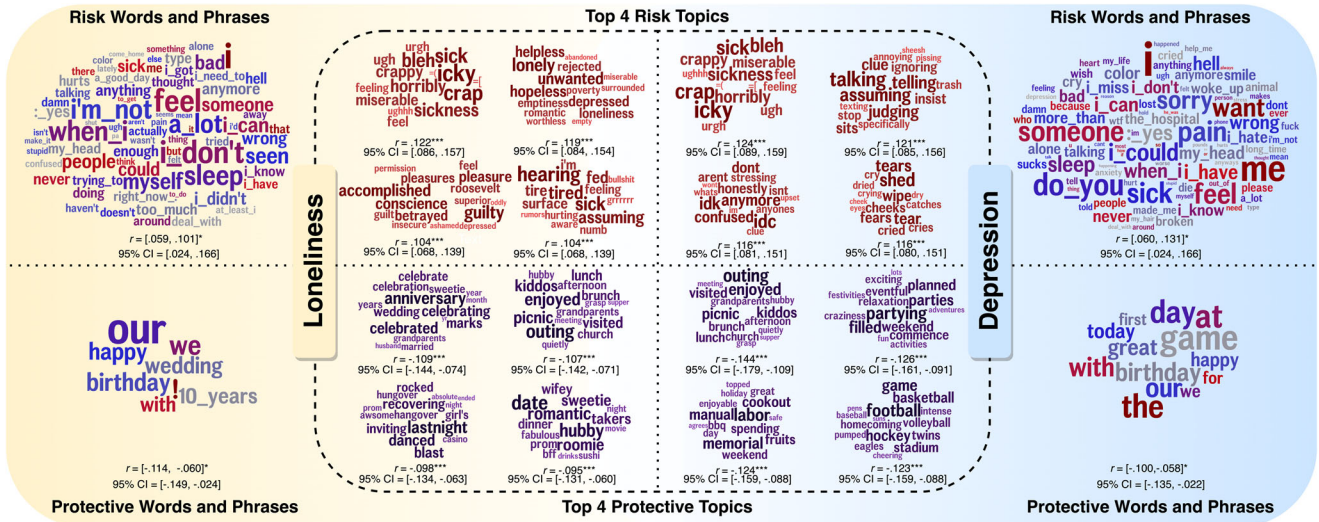
\*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$ .

Similar results were observed in open-vocabulary analyses (Fig. 1). Specifically, higher levels of loneliness and depression were both correlated with more language about negative emotions (LIWC: anxiety, anger, sad; words and phrases: 'feel', 'bad'; topics about negative feelings), sickness and pain (LIWC: health; words and phrases: 'pain', 'sick'; topics about pain and sickness), cognitive process (LIWC: differentiation, tentative, insight, cause; words and phrases: 'thought', 'know'; topics about cognition, observation, judgment), use of first-person singular pronouns (LIWC: I; words and phrases: 'I\_don't', 'I\_could'; topics about low self-worth), and other themes such as sleep, present temporal orientation, risk, death, interrogatives, and fillers.

Lower levels of loneliness and depression were associated with more linguistic markers of social relationships and activities (LIWC: we, affiliation; words and phrases: 'our', 'we', 'birthday', 'wedding', 'game'; topics about social relationships, social gatherings and activities). Words and phrases and the top four topics that were significantly associated with depression and loneliness are shown in Fig. 1, and the top 15 topics correlated with loneliness and depression with and without controlling for another are shown in Supplementary Table 4 and Supplementary Table 5, respectively.

### Linguistic markers unique to loneliness

After controlling for depression in addition to age and gender, there were two significant changes. First, controlling for depression made several linguistic markers of loneliness nonsignificant. That is, linguistic markers indicative of negative emotions, sickness and pain, first-person singular pronouns, and sleep were no longer correlated with loneliness. Second, after controlling for depression, correlations between loneliness and linguistic features reflecting social relationships (e.g., LIWC: we; topics about romantic relationships) remain statistically significant. Table 2 shows associations of top LIWC categories with loneliness before and after adding depression as an additional control. The top 15 correlated topics that were unique to loneliness and depression have been categorized and labeled by two psychologists and are presented with one example under each label, respectively, to showcase the topic contents in Fig. 2 (see the full list of top topics correlated with depression and loneliness with and without



**Fig. 1 Loneliness and depression have large overlaps in language markers, illustrated by words and phrases and top four Facebook topics.** All correlations in the table are controlled for age and gender. The collections of words and phrases results are present with a minimum significant level at  $p < .05$ . Risk = correlations with  $r > 0$ , protective = correlations with  $r < 0$ .  $r$  = Pearson's correlation coefficient, the range of  $r$  values: [min  $r$ -value, max  $r$ -value]. 95% CI 95% confidence interval: [lower bound, upper bound]. \*\*\* $p < .001$ , \*\* $p < .01$ . \* $p < .05$ . All  $p$ -values were corrected using the Benjamini–Hochberg False Discovery Rate correction. The font size of the word in each word cloud represents the correlation strength, the larger the size the stronger the correlation was.

controlling for around in Supplementary Table 4 and Supplementary Table 5).

Linguistic markers associated with a higher level of loneliness in cognition, thinking, reasoning, and observations remained after adding depression as an additional control, including LIWC categories of cognitive process (differentiations, tentative language, insight), LIWC function words (common adverbs, auxiliary verbs), the 1-gram word 'of' ( $r = 0.070$ , 95% confidence interval: [0.034, 0.106],  $p = 0.024$ ), as well as topics about unusualness (e.g., 'odd', 'unusual', 'strange') and observations (e.g., 'checked', 'sounded', 'mistaken'), writing and reading (e.g., 'poems', 'journal', 'writer'), and mixed emotions with cognitive components (e.g., 'amusing', 'oddy', 'disappointing'). See detailed statistics in Table 2 and Fig. 2.

Protective linguistic markers against loneliness described social relationships, including LIWC categories of 1st person plural, affiliation, family, and friends, the 1-gram word 'our' ( $r = -0.069$ , 95% confidence interval: [-0.104, -0.033],  $p = 0.024$ ), and topics about romantic relationships (e.g., 'husband', 'wife', 'wifey'), and (interactions with) family members and friends (e.g., 'hugged', 'sister', 'wiped'). These linguistic features are different from those associated with depression after controlling for loneliness (Fig. 2). As it is beyond the scope of this study to describe unique factors of depression after controlling for loneliness, we present the unique topics associated with depression in Fig. 2 without further discussion.

### Facebook language predicts loneliness and depression

The predictive power of Facebook language in predicting loneliness is shown in Table 3. We found that language features outperformed age and gender in predicting loneliness and depression, but the predictive performance of both demographics and Facebook posts is lower for loneliness ( $r = 0.133$ , using Age and Gender;  $r = 0.201$ , using BERT) as compared to depression ( $r = 0.253$  using Age and Gender;  $r = 0.312$ , using BERT).

## DISCUSSION

Our study has three main findings. Firstly, language markers associated with loneliness and depression overlap substantially

(over 50% of significant linguistic markers of loneliness were also correlated with depression). Secondly, the risk for loneliness has unique linguistic markers that reflect cognitive processing of environmental information, self-oriented cognitive activities (e.g., reading and writing), and mixed emotions with cognitive components. As may be expected, linguistic references to social relationships show protective associations with loneliness. Thirdly, the variance accounted for by predictive models based on demographics and the language in Facebook posts is lower for loneliness than depression.

Loneliness shares many features with depression, as measured by surveys and language use. Our observed moderate positive correlation between loneliness and depression ( $r = 0.554$ ) replicates other studies<sup>5,11</sup>. The linguistic markers positively correlated with both include negative emotions, sickness, and pain, as well as first-person singular pronouns. Linguistic markers negatively correlated with both include friends, family, and social gatherings. These correlational patterns are in line with the literature on the risky and protective linguistic markers of depression<sup>27,33,34,50</sup> and loneliness<sup>32,51,52</sup>, and the psychological constructs (e.g., worry) associated with both<sup>19</sup>. This also supports the plausible effectiveness of past interventions to reduce loneliness by asking patients to focus on reducing their internal negative thoughts<sup>20</sup>.

The significant correlation between loneliness and age became insignificant when controlling for depression, but the correlation between depression and age remained effectively unchanged when controlling for loneliness. This suggests that depression may be the third variable driving the relationship between age and loneliness. Given the high prevalence of loneliness among the elderly with depression<sup>4,8</sup>, this confound should be taken into account in future work.

We observed similar patterns in language: controlling for depression rendered insignificant the majority of linguistic features correlated with loneliness, but controlling for loneliness did not impact the linguistic features correlated with depression. Because we found large overlaps between the language of loneliness and depression, to understand the shared and unique linguistic features, we looked at partial language correlations of one construct controlling for the other. While this approach foregrounds the unique language patterns associated with either construct, as has been observed by other personality researchers,

**Table 2.** LIWC categories associated with loneliness with and without controlling for depression.

LIWC categories			Controlled variables					
			Age, Gender			Age, Gender, Depression		
Super category	Subcategory	Example words	<i>r</i>	95% CI		<i>r</i>	95% CI	
				L	U		L	U
<i>Risk factors</i>								
Affective processes	Anxiety	fear, worry, afraid	0.096***	0.060	0.131			
	Sadness	miss, lost, sad, sorry	0.090***	0.054	0.125			
	Anger	hate, hell, stupid	0.087***	0.051	0.122			
Cognitive processes	Differentiation	but, not, if, or,	0.107***	0.071	0.142	0.061**	0.025	0.097
	Tentative	if, or, some, any, hope, may	0.104***	0.068	0.139	0.055**	0.019	0.090
	Insight	know, think, feel, find, believe	0.095***	0.059	0.130	0.061**	0.026	0.097
Total function words	Common adverbs	so, just, when	0.103***	0.068	0.138	0.057**	0.022	0.093
	Negations	not, no, don't, can't	0.103***	0.067	0.138			
	Auxiliary verbs	is, have, be, are, was	0.093***	0.057	0.128	0.055**	0.020	0.091
Total pronouns	1st person singular	i, my, me, I'm, I've	0.101***	0.065	0.136			
	Impersonal pronouns	this, it, that	0.082***	0.046	0.118			
Other grammar	Common verbs	is, have, be, are, was, will, get, do	0.086***	0.051	0.122			
Time orientations	Present focus	is, have, be, are, get	0.075***	0.039	0.110			
Personal concerns	Death	dead, died, die, war	0.075***	0.039	0.110			
<i>Protective factors</i>								
Total pronouns	1st person plural	we, our, us, let's	-0.122***	-0.157	-0.087	-0.070***	-0.106	-0.035
Drives	Affiliation	we, love, our, friends	-0.113***	-0.148	-0.078	-0.079***	-0.114	-0.043
Social processes	Family	family, baby, mom	-0.078***	-0.113	-0.042	-0.058**	-0.094	-0.022
	Friends	friends, friend, date				-0.053**	-0.088	-0.017
Affective processes	Positive emotion	love, good, happy	-0.060**	-0.095	-0.024			
Biological processes	Ingestion	sweat, eat, water	-0.055**	-0.091	-0.019			

Top LIWC significant risk and all protective categories are present in the table. All correlations in the table are significant at  $p < 0.01$  level, insignificant correlations are left blank in the table. Risk factors are those with  $r > 0$ ; protective factors are those with  $r < 0$ .  $r$  = Pearson's correlation coefficient. All  $p$ -values were corrected using the Benjamini-Hochberg False Discovery Rate correction.

LIWC Linguistic Inquiry and Word Count, English 2015 category, 95% CI 95% confidence interval, L lower, U upper.

\*\*\* $p < 0.001$ , \*\* $p < 0.01$ .

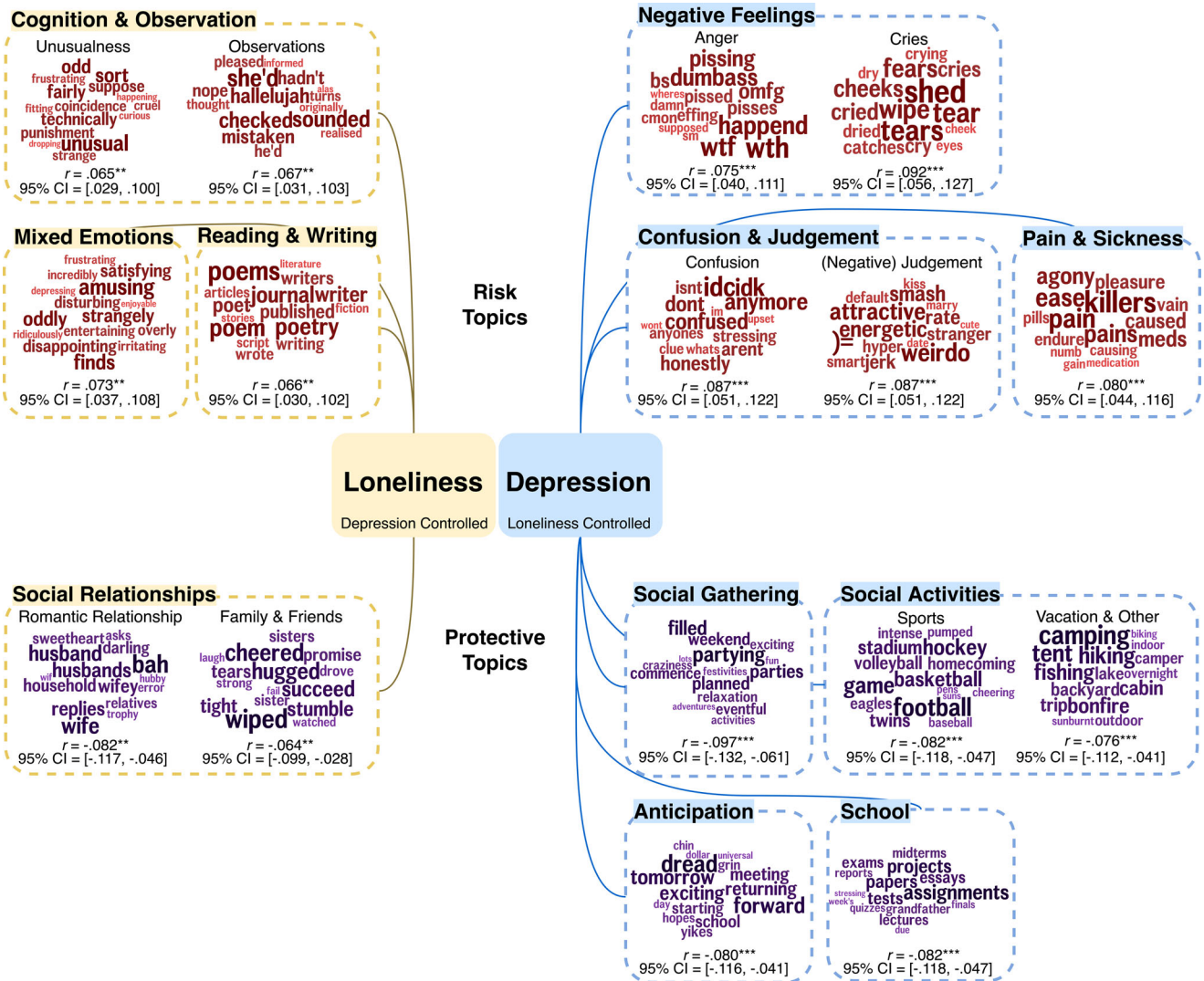
we also note that the interpretation of a psychological construct may become difficult once an overlapping, closely related psychological construct is partialled out—as is the case here<sup>53</sup>. Effect sizes are also reduced<sup>54</sup>.

It is worth noting that linguistic markers of loneliness reported in past research<sup>32,51,52</sup> were found to be also shared by depression in the present study, which further indicates that the impact of depression should not be overlooked in investigating loneliness. But our findings do not suggest that we should treat loneliness merely as a facet of depression just because depression shares many common features and heavily impacts loneliness. Instead, and more importantly, we found loneliness has unique manifestations in language use, which could be targets of personalized treatments.

We found that greater loneliness (unlike depression) is related to more references to greater cognitive reasoning and information processing of surroundings; while less loneliness is related to references to social relationships, after controlling for depression. Loneliness was linked to cognitive and information processing in prior cross-sectional and longitudinal surveys and neuroimaging studies<sup>24</sup>, serving as a protective mechanism against unpleasant social interaction and threats<sup>25</sup>. Lonely individuals are more likely to view the social world as threatening and pay more attention to and generate interpretations of the social environment, which may contribute to more biased expectations and attributions, and greater cognitive load in the processing and observation of

social surroundings<sup>55,56</sup>. This potential increase in cognitive load may be supported by the correlation between LIWC cognitive process (e.g., insight and tentative categories) and loneliness. Previous findings in psycholinguistics suggest that the use of insight words reflects active processes of cognitive appraisal, and the use of tentative language reflects uncertainty<sup>57</sup>. A more recent study testing the relationship between cognitive processes and the use of cognitive linguistic markers from LIWC also suggests that higher cognitive load is associated with greater use of cognitive linguistic markers<sup>58</sup>.

We also observed that loneliness is negatively correlated with language about close relationships (i.e., romantic relationships and friends and family members) and positive social interactions related to these relationships (e.g., hugging). Although positive celebrations and gatherings could reduce both depression and loneliness, loneliness is uniquely negatively associated with social relationships rather than activities and celebrations, indicating that loneliness might be particularly driven by the lack of personal and family experiences. The above findings also mean that loneliness could be reduced through more positive social experiences and the treatment of maladaptive social cognitions<sup>56</sup>. Past findings on treatment have shown that the positive social interactions and relationships an individual experiences can shape their social expectations and motivations, which has been linked to lower levels of loneliness<sup>24</sup>; interventions designed to address



**Fig. 2 Unique Facebook topics predicting loneliness versus depression.** Distinct Latent Dirichlet Allocation Topics were extracted and correlated with depression and loneliness scores, controlling for loneliness and depression scores in addition to age and gender, respectively. The top 15 correlated topics were categorized and labeled by two psychologists and presented with one example under each label to showcase the topic contents. The font size of the word in each word cloud represents the correlation strength, the larger the size the stronger the correlation was. Risk = correlations with  $r > 0$ , protective = correlations with  $r < 0$ .  $r$  = Pearson's correlation coefficient. 95% CI 95% confidence interval: [lower bound, upper bound]. All correlations in the table are significant at the  $p < .01$  level.  $***p < .001$ ,  $**p < .01$ . All  $p$ -values were corrected using the Benjamini–Hochberg False Discovery Rate correction.

maladaptive social cognition have been found to be most effective in reducing loneliness<sup>20</sup>.

Our findings provide consistent evidence from both closed- and open-vocabulary language analyses to support the above hypotheses, reflecting a “head versus heart” difference between loneliness and depression. That is, loneliness is positively associated with over-attention to the environment, including watchfulness towards changes, as well as mental preoccupations and reasoning, and self-directed cognitive activities like reading and writing (in the “head”). Depression, on the other hand, focuses on negative emotions, pain perception, and emotionally-focused rumination (from the “heart”).

The current study has limitations. First, the current study is correlational; the causal relationship between loneliness, depression, and linguistic features cannot be established. Second, our findings are based on a sample of adults in the United States; findings cannot easily be extended to teenagers or older adults—the two populations that most suffer from loneliness. Third, the current paper used a three-item scale for the

assessment of loneliness, though with good reliability, which limits the power and accuracy in predicting and measuring loneliness. Future studies should utilize more comprehensive assessments (e.g., the standard UCLA 20-item Loneliness scale<sup>42</sup> and other assessments besides self-report scales, including interviews) to evaluate the levels of loneliness. Additionally, the current study only conducted a single-time assessment of depression and loneliness yet included all of the participants' Facebook postings to extract linguistic features, which covered a wide period. The inclusion of entire timeline of Facebook posts enhances the data quantity and has been shown in the literature to produce fine-grained linguistic features, but did not foreground signal reflecting the episodic nature of depression and loneliness. Future studies should evaluate the language associated with changes in depression and loneliness over time.

When comparing the effectiveness of the current interventions to reduce loneliness, a meta-analysis<sup>20</sup> found that social cognitive training interventions in social groups are more effective than other intervention types (i.e., enhancing social skills, community-based

**Table 3.** Predictive power of linguistic models in the current study.

Features	Loneliness		Depression	
	<i>r</i>	MAE	<i>r</i>	MAE
Demographics				
Age + gender	0.133	1.90	0.253	4.63
Language				
LIWC	0.190	1.87	0.283	4.58
N-grams	0.155	1.89	0.209	4.71
Topics	0.169	1.88	0.274	4.56
BERT	0.201	1.87	0.312	4.54
Demographics + Language				
Age + gender + BERT	0.204	1.87	0.321	4.52
Age + gender + LIWC + N-gram + topics + BERT	0.204	1.87	0.312	4.55

*r* = Pearson's correlation coefficient. N-grams = words and phrases (1–3-grams). Topics = clusters of co-occurring words extracted using the Latent Dirichlet Allocation method.  
MAE Mean Absolute Error, LIWC Linguistic Inquiry and Word Count, English 2015 category. BERT Bidirectional Encoder Representations from Transformers.

groups). Our findings also provide evidence to support this. Reflected by language use, greater loneliness is linked to preoccupation with processing environmental information and self-oriented cognitive activities. Therefore, future interventions should perhaps consider targeting clients' perceptions, reasoning, cognitions, and relationships, especially in the context of the social environment. Potential training to reduce loneliness could focus on changing clients' cognitive style in understanding surroundings, shifting cognitive focus from changes to the regularities in the environment, reducing self-directed activities like reading and writing, and strengthening connections to close relationships. In addition, considering the large overlap between loneliness and depression, future interventions should also consider noticing and treating clients' other affective distress to better reduce loneliness.

#### DATA AVAILABILITY

De-identified data necessary to reproduce the results contained in the document is available upon request from the corresponding authors. We will not, however, share individual-level Facebook data as it contains potentially identifying information about participants enrolled in the study.

#### CODE AVAILABILITY

The code used in the present paper is now available on GitHub (<https://github.com/hellotingting/language-of-loneliness-wvbp>).

Received: 9 June 2022; Accepted: 12 September 2022;

Published online: 18 October 2022

#### REFERENCES

- Heinrich, L. M. & Gullone, E. The clinical significance of loneliness: a literature review. *Clin. Psychol. Rev.* **26**, 695–718 (2006).
- Hawkey, L. C. & Cacioppo, J. T. Loneliness matters: a theoretical and empirical review of consequences and mechanisms. *Ann. Behav. Med.* **40**, 218–227 (2010).
- Lim, M. H., Holt-Lunstad, J. & Badcock, J. C. Loneliness: contemporary insights into causes, correlates, and consequences. *Soc. Psychiatry Psychiatr. Epidemiol.* **55**, 789–791 (2020).
- Victor, C. R. & Yang, K. The prevalence of loneliness among adults: a case study of the United Kingdom. *J. Psychol.* **146**, 85–104 (2012).
- Beutel, M. E. et al. Loneliness in the general population: prevalence, determinants and relations to mental health. *BMC Psychiatry* **17**, 97 (2017).
- Cacioppo, J. T. & Cacioppo, S. The growing problem of loneliness. *Lancet* **391**, 426 (2018).
- Cacioppo, J. T. et al. Loneliness and health: potential mechanisms. *Psychosom. Med.* **64**, 407–417 (2002).
- Ebesutani, C. et al. The role of loneliness in the relationship between anxiety and depression in clinical and school-based youth. *Psychol. Schools* **52**, 223–234 (2015).
- Erzen, E. & Çikrikci, Ö. The effect of loneliness on depression: a meta-analysis. *Int. J. Soc. Psychiatry* **64**, 427–435 (2018).
- Palgi, Y. et al. The loneliness pandemic: loneliness and other concomitants of depression, anxiety and their comorbidity during the COVID-19 outbreak. *J. Affect. Disord.* **275**, 109 (2020).
- Russell, D., Cutrona, C. E., Rose, J. & Yurko, K. Social and emotional loneliness: an examination of Weiss's typology of loneliness. *J. Person. Soc. Psychol.* **46**, 1313 (1984).
- Weeks, D. G., Michela, J. L., Peplau, L. A. & Bragg, M. E. Relation between loneliness and depression: a structural equation analysis. *J. Person. Soc. Psychol.* **39**, 1238 (1980).
- Mahon, N. E., Yarcheski, A., Yarcheski, T. J., Cannella, B. L. & Hanks, M. M. A meta-analytic study of predictors for loneliness during adolescence. *Nursing Res.* **55**, 308–315 (2006).
- Beck, A. T. & Alford, B. A. *Depression: Causes and Treatment*. (University of Pennsylvania Press, 2009).
- Killgore, W. D., Cloonan, S. A., Taylor, E. C. & Dailey, N. S. Loneliness: a signature mental health concern in the era of COVID-19. *Psychiatry Res.* **290**, 113117 (2020).
- Cacioppo, J. T., Hughes, M. E., Waite, L. J., Hawkey, L. C. & Thisted, R. A. Loneliness as a specific risk factor for depressive symptoms: cross-sectional and longitudinal analyses. *Psychol. Aging* **21**, 140 (2006).
- Cacioppo, J. T., Hawkey, L. C. & Thisted, R. A. Perceived social isolation makes me sad: 5-year cross-lagged analyses of loneliness and depressive symptomatology in the Chicago Health, Aging, and Social Relations Study. *Psychol. Aging* **25**, 453 (2010).
- Burger, J. et al. Bereavement or breakup: differences in networks of depression. *J. Affect. Disord.* **267**, 1–8 (2020).
- Anyan, F., Morote, R. & Hjemdal, O. Prospective relations between loneliness in different relationships, metacognitive beliefs, worry and common mental health problems. *Mental Health Prevent.* **19**, 200186 (2020).
- Masi, C. M., Chen, H.-Y., Hawkey, L. C. & Cacioppo, J. T. A meta-analysis of interventions to reduce loneliness. *Person. Soc. Psychol. Rev.* **15**, 219–266 (2011).
- Vanhals, J., Luyckx, K., Raes, F. & Goossens, L. Loneliness and depressive symptoms: the mediating and moderating role of uncontrollable ruminative thoughts. *J. Psychol.* **146**, 259–276 (2012).
- Danneel, S. et al. Loneliness, social anxiety, and depressive symptoms in adolescence: examining their distinctiveness through factor analysis. *J. Child Family Studies* **28**, 1326–1336 (2019).
- Lasgaard, M., Goossens, L. & Elklit, A. Loneliness, depressive symptomatology, and suicide ideation in adolescence: Cross-sectional and longitudinal analyses. *J. Abnormal Child Psychol.* **39**, 137–150 (2011).
- Cacioppo, J. T. & Hawkey, L. C. Perceived social isolation and cognition. *Trends Cogn. Sci.* **13**, 447–454 (2009).
- Cacioppo, J. T. et al. Loneliness within a nomological net: an evolutionary perspective. *J. Res. Person.* **40**, 1054–1085 (2006).
- Akhter-Khan, S. C. & Au, R. Why loneliness interventions are unsuccessful: a call for precision health. *Adv. Geriatr. Med. Res.* **2**, e200016 (2020).
- Eichstaedt, J. C. et al. Facebook language predicts depression in medical records. *Proc. Natl Acad. Sci. USA* **115**, 11203–11208 (2018).
- Ophir, Y., Tikochinski, R., Brunstein Klomek, A. & Reichart, R. The hitchhiker's guide to computational linguistics in suicide prevention. *Clin. Psychol. Sci.* **10**, 212–235 (2022).
- Eichstaedt, J. C. et al. Closed- and open-vocabulary approaches to text analysis: A review, quantitative comparison, and recommendations. *Psychol. Method.* **26**, 398–427 (2020).
- Kern, M. L. et al. Gaining insights from social media language: methodologies and challenges. *Psychol. Methods* **21**, 507 (2016).
- Schwartz, H. A. et al. Personality, gender, and age in the language of social media: the open-vocabulary approach. *PLoS ONE* **8**, e73791 (2013).
- Guntuku, S. C. et al. Studying expressions of loneliness in individuals using twitter: an observational study. *BMJ Open* **9**, e030355 (2019).
- Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H. & Eichstaedt, J. C. Detecting depression and mental illness on social media: an integrative review. *Curr. Opin. Behav. Sci.* **18**, 43–49 (2017).

34. Liu, T. et al. Cross-platform difference in facebook and text messages language use: illustrated by depression diagnosis. *arXiv* <https://arxiv.org/abs/2202.01802> (2022).
35. De Choudhury, M., Gamon, M., Counts, S. & Horvitz, E. Predicting depression via social media. In *Seventh international AAAI conference on weblogs and social media*. **7** (2013).
36. Perrin, A. & Anderson, M. Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018, Vol. 48, 573–585 <https://www.pewresearch.org/fact-tank/2019/2004/2010/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/> (Pew Research Center, 2019).
37. Liu, T. et al. Linguistic predictors from Facebook postings of substance use disorder treatment retention versus discontinuation. *Am. J. Drug Alcohol Abuse* **48**, 573–585 (2022).
38. Ophir, Y., Tikochinski, R., Asterhan, C. S., Sisso, I. & Reichart, R. Deep neural networks detect suicide risk from textual facebook posts. *Sci. Rep.* **10**, 1–10 (2020).
39. Hughes, M. E., Waite, L. J., Hawkey, L. C. & Cacioppo, J. T. A short scale for measuring loneliness in large surveys: results from two population-based studies. *Res. Aging* **26**, 655–672 (2004).
40. Kroenke, K., Spitzer, R. L. & Williams, J. B. The PHQ-9: validity of a brief depression severity measure. *J. Gen. Intern. Med.* **16**, 606–613 (2001).
41. Pennebaker, J. W., Boyd, R. L., Jordan, K. & Blackburn, K. The development and psychometric properties of LIWC2015. (University of Texas, 2015).
42. Russell, D., Peplau, L. A. & Cutrona, C. E. The revised UCLA Loneliness Scale: concurrent and discriminant validity evidence. *J. Person. Soc. Psychol.* **39**, 472 (1980).
43. Donovan, N. J. et al. Association of higher cortical amyloid burden with loneliness in cognitively normal older adults. *JAMA Psychiatry* **73**, 1230–1237 (2016).
44. Michinov, E. & Michinov, N. Stay at home! When personality profiles influence mental health and creativity during the COVID-19 lockdown. *Curr. Psychol.* 1–12 (2021).
45. Cornwell, E. Y. & Waite, L. J. Social disconnectedness, perceived isolation, and health among older adults. *J. Health Soc. Behav.* **50**, 31–48 (2009).
46. Kroenke, K. & Spitzer, R. The PHQ-9: a new depression diagnostic and severity measure. *Psychiatric Ann.* **32**, 509–521 (2002).
47. Arroll, B. et al. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. *Ann. Family Med.* **8**, 348–353 (2010).
48. Blei, D. M., Ng, A. Y. & Jordan, M. I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **3**, 993–1022 (2003).
49. Liu, T., Ackerman, J. M. & Preston, S. D. Dissociating compulsive washing and hoarding tendencies through differences in comorbidities and the content of concerns. *J. Behav. Cogn. Ther.* **31**, 291–308 (2021).
50. Edwards, T. M. & Holtzman, N. S. A meta-analysis of correlations between depression and first person singular pronoun use. *J. Res. Person.* **68**, 63–68 (2017).
51. Badal, V. D. et al. Do Words matter? Detecting social isolation and loneliness in older adults using natural language processing. *Front. Psychiatry* **12**, 1–12 (2021).
52. Andy, A. Studying how individuals who express the feeling of loneliness in an online loneliness forum communicate in a nonloneliness forum: observational study. *JMIR Form. Res.* **5**, e28738 (2021).
53. Lynam, D. R., Hoyle, R. H. & Newman, J. P. The perils of partialling: cautionary tales from aggression and psychopathy. *Assessment* **13**, 328–341 (2006).
54. Paulhus, D. L., Robins, R. W., Trzesniewski, K. H. & Tracy, J. L. Two replicable suppressor situations in personality research. *Multivariate Behav. Res.* **39**, 303–328 (2004).
55. Cacioppo, J. T. & Hawkey, L. C. People Thinking About People: The Vicious Cycle of Being a Social Outcast in One's Own Mind. In *The social outcast: Ostracism, social exclusion, rejection, and bullying* (Eds. Williams, K. D., Forgas, J. P. & von Hippel, W.) 91–108 (Psychology Press, 2005).
56. Cacioppo, S., Grippo, A. J., London, S., Goossens, L. & Cacioppo, J. T. Loneliness: clinical import and interventions. *Perspect. Psychol. Sci.* **10**, 238–249 (2015).
57. Tausczik, Y. R. & Pennebaker, J. W. The psychological meaning of words: LIWC and computerized text analysis methods. *J. Lang. Soc. Psychol.* **29**, 24–54 (2010).
58. Konopasky, A., Durning, S. J., Artino, A. R., Ramani, D. & Battista, A. The linguistic effects of context specificity: exploring affect, cognitive processing, and agency in physicians' think-aloud reflections. *Diagnosis* **7**, 273–280 (2020).

## ACKNOWLEDGEMENTS

This study was partially funded by the Intramural Research Program of the National Institutes of Health (NIH), National Institute on Drug Abuse (NIDA). T.L. and S.C.G. share the corresponding authorship. The co-corresponding authors, T.L. and S.C.G., had full access to all the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis.

## AUTHOR CONTRIBUTIONS

S.C.G., L.H.U., L.T. and G.S. originated the study and acquisition of the data. S.C.G. and G.S. developed methods. T.L. did critical analyses, created and revised the figures and tables, interpreted the findings, drafted the manuscript, and edited the manuscript. S.C.G., L.H.U., G.S., J.C.E., B.C., L.T. and K.Y. assisted with the interpretation of the findings and contributed to the writing/revision of the manuscript.

## FUNDING

Open Access funding provided by the National Institutes of Health (NIH).

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s44184-022-00014-7>.

**Correspondence** and requests for materials should be addressed to Tingting Liu or Sharath Chandra Guntuku.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as 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. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2022