



Anxiety and Depression are Associated with More Distorted Thinking on Social Media: A Longitudinal Multi-Method Study

Lauren A. Rutter^{1,3} · Andy Edinger^{3,4} · Lorenzo Lorenzo-Luaces^{1,3} · Marijn ten Thij⁵ · Danny Valdez^{2,3} · Johan Bollen^{3,4}

Accepted: 29 January 2025
© The Author(s) 2025

Abstract

Background Depression and anxiety are associated with patterns of negative thinking that can be targeted through cognitive restructuring as a part of cognitive therapy (CT) or cognitive behavioral therapy (CBT). Our team has created a set of cognitive distortion schemata (CDS) n-grams based on theories underlying CT to measure the linguistic markers that indicate cognitive vulnerability to depression. These CDS were specifically designed to examine online language. Our prior work supports a relationship between CDS and a diagnosis of depression, but less is known about the relationship between online language, CDS, and anxiety. The current study measures if CDS can be detected in people who report anxiety symptoms, and whether CDS increase with symptom severity.

Methods 1,377 participants were recruited from a study assessing social media use and mental health symptoms, the Studies of Online Cohorts of Internalizing Symptoms and Language (SOCIAL). From this, 804 timelines were harvested, and after removing missing data and bots, our final sample was 537 respondents who posted 999,859 tweets. This is a longitudinal, multi-method design, using surveys and text-based analysis of social media timelines. We used bootstrap resampling to compare differences in CDS prevalence in anxious and depressed participants.

Results CDS can be observed in anxiety disorders, significantly increase as a function of anxiety symptom severity, and are related to depression and anxiety comorbidity.

Conclusions Using behavioral, affective, and cognitive indicators of distorted thinking from social media may yield new insight into the trajectories of depression and anxiety. This work has implications for the future of CT/CBT and other online interventions that target distorted thinking styles.

✉ Lauren A. Rutter
larutter@iu.edu

Andy Edinger
aedinge@iu.edu

Lorenzo Lorenzo-Luaces
lolorenz@iu.edu

Marijn ten Thij
m.c.tenthij@tilburguniversity.edu

Danny Valdez
danvald@iu.edu

Johan Bollen
jbollen@iu.edu

¹ Department of Psychological and Brain Sciences, Indiana University Bloomington, Bloomington, USA

² Department of Applied Health Science, Indiana University Bloomington, Bloomington, USA

³ Center for Social and Biomedical Complexity, Indiana University Bloomington, Bloomington, USA

⁴ Cognitive Science Program, Indiana University Bloomington, Bloomington, USA

⁵ Department of Cognitive Science and Artificial Intelligence, Tilburg University, Tilburg, Netherlands

Introduction

Anxiety and mood disorders, also known as internalizing disorders, are among the most common mental disorders and leading contributors to the burden of disability worldwide (Greenberg et al., 1999, 2015; Konnopka & König, 2020). Internalizing disorders are heterogeneous in their symptoms (Fried et al., 2020; Lorenzo-Luaces et al., 2020, 2021), prognosis (Lorenzo-Luaces, 2015), and impact multiple bodily systems including circadian cycles (Ten Thij et al., 2020), emotion processing (Rutter et al., 2022; Rutter, Norton et al., 2020; Rutter, Passell, Rutter et al., 2019, 2020a, b, 2024), hormones (Granger et al., 1994; Young & Korszun, 2010), physical health (Berkman et al., 1986; Rutter et al., 2021), and even language (Bathina et al., 2021; Bollen et al., 2021). Regarding language, prior work has shown that individuals with symptoms of depression more frequently use first-person pronouns, absolutist language, and terms that describe negative emotions (Al-Mosaiwi & Johnstone, 2018; Bernard et al., 2016; Rude et al., 2004; Tackman et al., 2019). These findings are consistent with research suggesting that internalizing disorders are characterized by a self-focus and negative affectivity. Individuals with a diagnosis of depression also display more distorted negative thinking in real life (Beck, 1963), and in their online communications (Bathina et al., 2021).

Aspects of distorted thinking manifest by distinctive word choices and sentence structure are often targeted directly in therapy. Indeed, one of the leading evidence-based treatment for internalizing disorders is cognitive behavioral therapy (CBT), which aims to restructure negative thinking and change associated behaviors (Lorenzo-Luaces et al., 2015, 2016). Investigating the language of individuals with internalizing disorders may assist in understanding this heterogeneous and complex disorder. A better understanding of language indicative of depression and related disorders could assist in early detection or more individualized treatments. The relationship between distorted thinking in anxiety and depression and the ways in which these distorted thinking styles contribute to and perpetuate symptoms is well-established (e.g., Beck & Clark, 1988; Eysenck et al., 2007). Cognitive distortions are central to creating and maintaining symptoms across internalizing disorders. The novelty of our approach is using social media data to explore cognitively distorted language. Social media provides a platform where researchers can learn more about the language of depression and anxiety, as it naturally occurs. In the United States, over 70% of individuals are on a social media platform. The insights gained from social media data can have direct impacts on treatment of internalizing disorders through the use of machine learning and AI tools that can detect depressive thinking styles and could be used for

the purposes of prevention and early intervention (Babu & Kanaga, 2022; Deshpande & Rao, 2017). This work has opportunities to contribute to the growing field of large language models (LLMs) in mental health (Lawrence et al., 2024). Moreover, understanding vulnerability to internalizing disorder is important because social media itself may be associated with internalizing symptoms (Rutter et al., 2021).

Grounded in cognitive therapy and Beck's cognitive theory of depression, we have previously demonstrated that cognitive distortion schemata (CDS), or the patterns of thought represented by sequences of words (n-grams), differ between a depressed sample and a random sample on Twitter (Bathina et al., 2021). We created the CDS among a team of clinical psychologists and experts in natural language processing (NLP). We employed a theory-driven approach in creating the CDS, derived to measure the linguistic markers that indicate cognitive vulnerability to depression. The CDS categories we developed are 12 widely accepted styles of distorted thinking, including all-or-nothing thinking, catastrophizing, fortune telling, and others. Previous results that show CDS are higher in depressed individuals compared to a random sample of adults while controlling for other elements of depressogenic language including first-person pronouns and more negative language (Holtzman et al., 2017; Tackman et al., 2019). In other words, our findings *could not be explained* by higher use of "I/me/my" or more words like "sad/lonely/depressed." Rather, differences can be more appropriately explained by language structure and patterns themselves. Indeed, we have also found evidence that this kind of language may be increasing over time (Bollen et al., 2021), which connects with concerns about the negative effects of social media.

While we have previously shown higher rates of CDS in depressed Twitter users compared to a random sample (Bathina et al., 2021; Bollen et al., 2021), we have not explored cognitive distortions online among anxious people. There is some recent evidence that natural language sentiment is an indicator of depression and anxiety symptoms (Kaźmierczak et al., 2024), but this work is not informed by cognitive theories of depression or cognitive therapy (CT). Kaźmierczak and colleagues' recent work uses positively (e.g., beautiful) and negatively (e.g., horrible) valenced words and match these to severity of depression and anxiety symptoms, while our work developed a lexicon based on Beck's cognitive distortions. While valence is useful to better understand the connections between word use and symptoms, the general lack of theory-driven language processing work is a major gap in the literature. A better understanding of naturalistic (casual, online) cognitively distorted language could provide better characterizations of internalizing disorders, which could be helpful in more precise treatment targets given the high comorbidity of anxiety

and depressive disorders (Lorenzo-Luaces et al., 2022; Rutter & Brown, 2015; Rutter et al., 2023). Moreover, while we have previously examined prevalence rates of CDS, we have not examined if their use is related to symptom severity on a continuum which comports with a modern understanding of depression and anxiety.

Social media data provides a unique and naturalistic way of understanding language and its relationship to psychopathology. Rather than using language transcripts from therapy or coding language for content and themes, big data approaches allow for better understanding of language use and thinking styles at the population-level. A better characterization of cognitively distorted language online can help inform large language models for mental health, which could be ultimately used to enhance interventions. We would expect that more severe symptoms of anxiety and depressive psychopathology is associated with greater use of cognitive distortions. Thus, the current study was centered on two main research questions. First, does CDS prevalence increase as a function of anxiety severity (RQ1)? Second, how is CDS prevalence related to anxiety and depression comorbidity (RQ2)? For both of these research questions, we proposed a hypothesis. First, we expected that the prevalence of cognitive distortions in online language (calculated based on proportion of tweets containing CDS n-grams/total tweets) could be observed in individuals with anxiety disorders. and we hypothesized that CDS prevalence would increase as anxiety symptom severity increased. Second, we hypothesized that we would observe a relationship between depression severity and CDS prevalence, with higher levels of depression and anxiety associated with the highest proportion of distorted thinking in online language.

Method

This study was approved by the institutional review board of Indiana University Bloomington (2002549202 and 2005948214). We aimed to recruit approximately 1000 online participants who were Twitter users between July 2020 and March 2021.

Participants

We recruited participants via Qualtrics panels for a study on “social media and mental health” between July 2020 and March 2021. Participants were purposefully selected to represent the United States demographic distributions of age, gender, race, and ethnicity and received financial compensation for their participation. Participants were part of the Survey Online Cohorts for Internalizing Symptoms and Language (SOCIAL), specifically SOCIAL-I, described in

prior work (Lorenzo-Luaces et al., 2022; Rutter et al., 2023). Participants ($N = 3,472$) were asked to provide their Twitter handle in addition to answering a variety of self-report questions addressing mental health. From the provided Twitter handles ($N = 1,377$), personal Twitter timelines were harvested ($N = 804$) and evaluated for being human vs. bot-like (see Lorenzo-Luaces et al., 2023 for more detailed analysis). Participants were removed from our analysis if they did not fully complete the survey or if we were not able to retrieve a unique Twitter timeline based on the provided Twitter handle for that participant ($N = 691$). Finally, we ran a bot-analysis using Botometer (Sayyadiharikandeh et al., 2020) on each of these profiles and removed the accounts that were deemed very likely to be automated. Based on these criteria, our final sample of individuals consists of 537 respondents, who posted 999,859 tweets in their timelines. Demographic information is included in Table 1.

Measures

Severity Measure for Generalized Anxiety Disorder

The severity measure for generalized anxiety disorder (GAD) consists of 10 items rated on a 5-point Likert scale ranging from 0 (“never”) to 4 (“all of the time”). The participants were asked to rate the frequency of worry and associated symptoms over the past 7 days. Total scores range from 0 to 40. Higher scores indicate a greater severity of GAD symptoms. The average score (total raw score/number of items answered) can be used as a proxy for GAD severity: 0 indicates none, 1 indicates mild, 2 indicates moderate, 3 indicates severe, and 4 indicates extreme anxiety. In our sample, Cronbach’s alpha was 0.92 for SOCIAL-I indicating excellent internal consistency.

Severity Measure for Depression

The severity measure for depression is also known as the Patient Health Questionnaire-9 (PHQ-9) (Kroenke et al., 2001) and consists of 9 items rated on a 4-point Likert scale. Items are rated based on symptoms over the last 7 days, and range from 0 (“not at all”) to 3 (“nearly every day”). The PHQ-9 has excellent psychometric properties. We scored the PHQ-9 to align with the severity scores of the GAD-10, described above, and based on established scoring. PHQ-9 scores correspond with an aggregate severity measure as follows: 0–4: none (0), 5–9: mild (1), 10–14: moderate (2), 15–19: severe (3), and 20–27: extreme (4). In our sample, Cronbach’s alpha was 0.87 for SOCIAL-I suggesting that this measure was internally consistent.

Table 1 Demographic information

Variable	Value
Age (years), mean (SD)	33.53(13.09)
Race, n(%)	
White	372(69.27)
Black or African American	64(11.92)
American Indian or Alaska Native	4(0.74)
Asian	30(5.59)
Hispanic or Latino/a	52(9.68)
Other	15(2.79)
Sex, n(%)	
Male	154(28.68)
Female	372(69.27)
Prefer not to say	1(0.19)
Gender, n(%)	
Male	158(29.42)
Female	361(67.23)
Other	1(0.19)
Non-binary, Genderqueer or agender	7(1.30)
Sexual orientation, n(%)	
Bisexual/pansexual	71(13.22)
Heterosexual/straight	419(78.03)
Homosexual/gay	22(4.10)
Other (e.g., queer, questioning, asexual)	11(2.05)
Prefer not to say	4(1.86)
Annual household income (US \$), n(%)	
Less than \$10,000	54(10.06)
\$10,000 to \$19,999	59(10.99)
\$20,000 to \$29,999	81(15.08)
\$30,000 to \$39,999	45(8.38)
\$40,000 to \$49,999	32(5.96)
\$50,000 to \$59,999	44(8.19)
\$60,000 to \$69,999	31(5.77)
\$70,000 to \$79,999	33(6.15)
\$80,000 to \$89,999	26(4.84)
\$90,000 to \$99,999	29(5.40)
\$100,000 to \$149,999	53(9.87)
\$150,000 or more	40(7.45)

Note 10 participants did not respond to the demographic portion of the survey, so total for each demographic category is 527 participants of total 537

Cognitive Distortion Schemata

Beck proposed the concept of cognitive distortions to characterize the thinking of individuals with depression (Beck, 1963). We drew on these latest lists of cognitive distortions, which consist of 12 cognitive distortions categories. We (L.A.R., L.L-L., and J.B.) iteratively designed a list of 241 CDS n-grams in consultation with CBT experts. Each CDS n-gram is each geared to express at least one type of cognitive distortion. The CDS were formulated to capture the minimal semantic building blocks of distorted thinking while avoiding expressions that are specific to depression-related topics, such as poor sleep or ongoing health issues.

Where possible, higher-order n-grams were chosen to capture as much of the semantic structure of one or more distorted schemata as possible (Bathina et al., 2021). For example, the 3-gram ‘everyone will believe’ captures both ‘overgeneralizing’ and ‘mindreading’. For more details on CDS construction, as well as a details on CDS n-grams and relevant grammatical features, see Bathina et al., 2021. While our initial work examined CDS across the 12 categories, the current project collapses all CDS categories into one, representing all types of distorted thinking. We did this for ease of analysis, interpretation, and because we did not make conceptual distinctions between the categories based on our hypotheses. To collapse CDS categories into one metric of CDS, we simply calculate the proportion of all tweets that contain any occurrence of a CDS n-gram for each participant in our study as a measure of CDS prevalence.

Data Analysis

We report descriptive statistics for the demographic variables of participants in SOCIAL-I. Then, we used bootstrap resampling to compare differences in CDS prevalence by anxiety and depression severity. We calculate the between-group CDS prevalence as the proportion of CDS-containing tweets produced by the subset of the re-sampled group with a given severity score. In other words, for each severity class C we have a corresponding set of tweets T_C , the set of all tweets produced by the individuals in C . Our CDS n-gram schemata is a function $F(t) \rightarrow \{0,1\}$ that maps each tweet t to 1 if it contains any CDS n-gram or 0 otherwise. The prevalence for each severity class C is thus calculated as. To establish accuracy of these prevalence calculations, we apply bootstrap re-sampling. Our bootstrap analysis comprised randomly re-sampling with replacement n individuals from our sample population, where n is the size of our sample. This bootstrap estimates are calculated by resampling repeated $B=10,000$ times, and the CDS prevalence recorded at each step. Our bootstrap analysis thus produces a distribution of CDS prevalence estimates such that overlap of inner 95 percentile interval of the distribution of prevalence estimates for one class with the median of another class is considered to be indicative of non-significant differences between severity groups. As noted above, there were 5 severity groups for GAD (none, mild, moderate, severe, extreme) for which we bootstrap the CDS prevalence across the individuals in the severity group (Fig. 1). However, in our data sample the extreme severity class comprised only 16 individuals. This sample size was too small to be considered valid for bootstrap analysis and so was excluded from the final analysis. We followed this same process to examine depression severity and CDS prevalence, again using

Fig. 1 Bootstrapped aggregate CDS prevalence distributions for each GAD-10 severity class. *Note* Each bar above with the sample size of the corresponding severity class. The colored box represents interquartile range, while the horizontal lines correspond to 95% CI. Our results show a trend of increasing CDS prevalence as severity increases, with pairwise significant differences denoted by braces

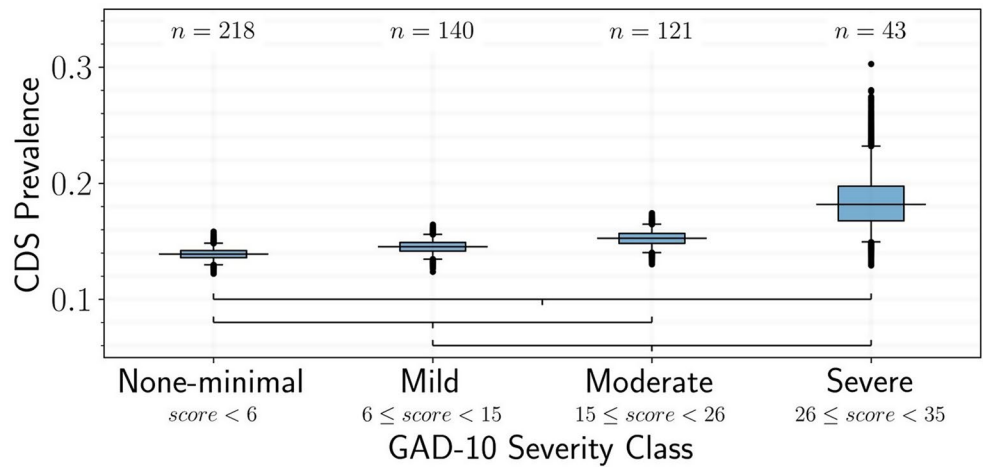


Fig. 2 Bootstrapped aggregate CDS prevalence distributions for each PHQ-9 severity class. *Note*. Each bar is annotated above with the median and 95% CI bounds in brackets. The colored box represents interquartile range, while the horizontal lines correspond to 95% CI. Pairwise significant differences denoted by braces

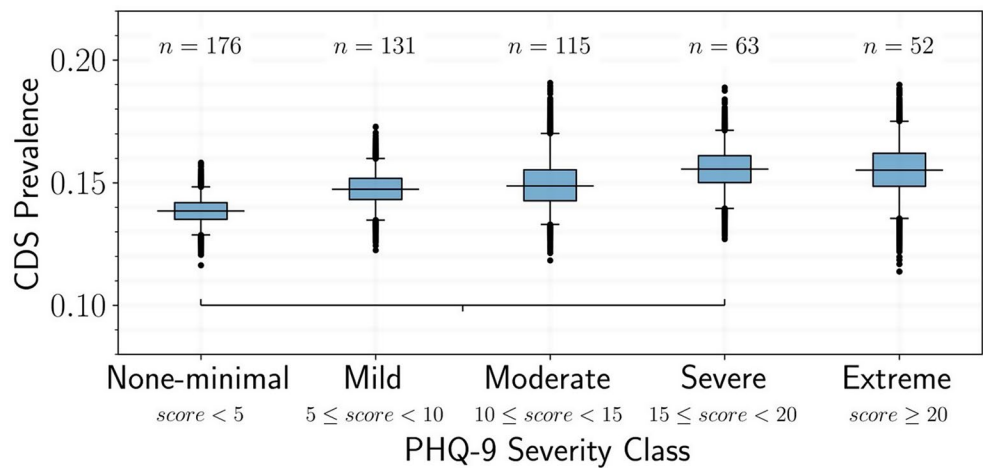
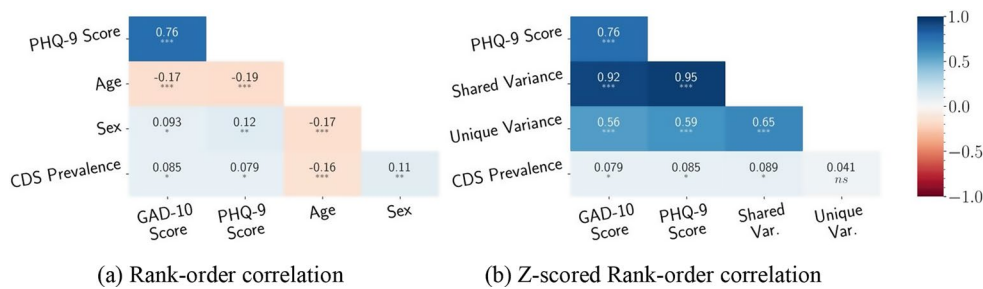


Fig. 3 Pairwise Spearman rank-order correlation coefficients between (a) GAD10, PHQ9 and confounding variables and (b) accounting for shared and unique variance between PHQ9 and GAD10. *Note*. Significance of results denoted below the coefficient value by **: p < 0.01 and *: p < 0.05



bootstrapping and examining 95% confidence interval overlaps (Fig. 2).

To further test the relationships between variables, we next conducted Spearman rank-order correlation between anxiety, depression, age, sex, and CDS prevalence (Fig. 3a). This correlation measure was selected due to concerns with respect to the underlying assumptions of aggregate Likert-scale questionnaire scores as linear measures of symptom severity. Rank-order correlation preserves ordering of severity without relying on the assumption that such questionnaire measures are linear metrics of underlying psychological traits (Liddell & Kruschke, 2018). We further examine the relationship between comorbidity of anxiety

and depression with CDS prevalence by computing Spearman rank-order correlations over the shared and unique variance (Hoza et al., 2005) between anxiety and depression (Fig. 3ab). Here, we calculate the shared variance between GAD-10 score and PHQ-9 score as $\frac{z_d - z_a}{2}$ and unique variance as $\frac{z_d + z_a}{2}$, where z_d is the z-scored PHQ-9 score and z_a is the z-scored GAD-10 score. These values thus account for the shared and marginal effects of depression and anxiety.

Results

Demographic information is shown in Table 1.

We also include Spearman rank-order correlation as shown in Fig. 3. Rank correlation demonstrates statistically significant relationships across all variables, although the effects are small, as has previously been reported when correlating NLP metrics with self-report (Holtzman et al., 2017). Anxiety (GAD-10) and depression (PHQ-9) scores were highly correlated with a correlation coefficient of 0.76, closely reflecting comorbidity rates between anxiety and depression. These severity scores were also correlated with CDS prevalence at a rate of 0.085 and 0.079, respectively. These correlation values are consistent with prior studies applying NLP to self-reported mental health data (DeSouza et al., 2021; Milintsevich et al., 2023).

To test our first hypothesis that CDS was related to anxiety symptom severity, we conducted regression. As shown in Fig. 1, there is a relationship between anxiety and online cognitive distortions which can be observed by increased prevalence of CDS, with individuals with no/minimal anxiety having the lowest CDS prevalence. There is an apparent increasing trend of CDS prevalence rising with GAD severity class, with moderate and severe GAD scores showing significantly greater prevalence than none-minimal and mild GAD. However, differences of none-minimal versus mild or moderate versus severe do not reach statistical significance. In general, we found support for our hypothesis, showing that CDS prevalence increases with increases in anxiety severity.

To test RQ2, we examined the relationship between CDS prevalence and depression severity. In examining CDS prevalence and depression severity, we see a linear increase in CDS prevalence by severity (Fig. 2). However, there is no significant difference between extreme levels of depression and minimal levels of depression. We attribute this to the lower number of individuals that we could include in these categories, as the bootstrapped intervals are the widest for this category.

Overall, our results supported our hypotheses. First, CDS increase as anxiety severity increases. Support of our second hypothesis that GAD and depression comorbidity would produce high levels of CDS is more mixed. As our rank-order correlation analysis shows a high correlation between GAD-10 and PHQ-9 scores (see Fig. 3), our cohort unfortunately does not lend itself to comorbidity analyses due to a low number of observations of individuals that score high on one of the questionnaires and low on the other. However, adding support to our prior work, we see CDS increase as depression severity increases for minimal through severe levels of depression (see Fig. 2).

Discussion

To the best of our knowledge, this is the first study to examine cognitive distortions in online language in mood and anxiety disorders using natural language processing (NLP) and a theory-based lexicon of n-grams. While some prior work using natural language processing has shown differences in language and thinking styles in depressed vs. random cohorts, less focus has been given to anxiety. This is a critical gap in our understanding of the thinking styles across internalizing disorders, especially considering that anxiety disorders increase the risk for mood disorders (Black et al., 2010; Duffy et al., 2013; Hong & Cheung, 2015), and that anxiety disorders alone are associated with a very high level of distress, impairment, and disability (Greenberg et al., 1999).

Our primary finding was that CDS increase as levels of generalized anxiety symptoms increase. We observed a similar trend with major depression symptoms as based on the PHQ-9: as symptoms of depression increased, prevalence of distorted thinking increased, controlling for age and sex covariates. While shared variance of anxiety and depression symptoms contribute to overall CDS prevalence, the size of the effect is slightly higher than unique variance, according to Spearman rank-order correlations (0.089=shared vs. 0.041=unique). Regarding the size of the effects we observed, there are many reasons why variance in CDS prevalence that is explained by severity is expected to be low. For example, there are many other variables that impact expression of CDS including time of year and the content of the post. Indeed, in a recent study examining emotion connectivity and symptoms of depression, sizes of the effect were similarly small (Kelley et al., 2023).

Based on the CDS construction, which was informed by Beck's original cognitive distortion categories (Beck, 1963), and expanded by 10 CBT experts in our team (Bathina et al., 2021), we provide validation of distorted thinking in anxiety disorders in online language. This work sheds additional light on the degree to which cognitively distorted and depressogenic language occurs colloquially in social media platforms. Rather than developing a classifier to detect anxiety and depression online, which is already being done by other research groups, our work seeks to better characterize distorted thinking online, naturalistically. Our CDS development and subsequent work based in CT is theory-driven, as opposed to purely data-driven NLP approaches. This work can be used to better understand how language is used in vulnerable individuals within social networks. Moreover, our work can be used in tandem with emerging work using LLMs applied to support mental health (Lawrence et al., 2024), and the growing body of literature applying NLP and machine learning to mental health (Le Glaz et al., 2021).

There is the potential for largescale societal relevance of our work: as individuals connect across the globe, they may use distorted language in specific ways within their social networks (Edinger et al., 2023; Tong et al., 2018). There may be elements of contagion (Al-Mosaiwi & Johnstone, 2018; Lee & Theokary, 2021) of distorted language within certain communities with depression and anxiety disorders, but this remains to be tested.

Despite the novelty of our research, using a theory-driven approach to capture distorted language, there are several limitations to consider. First, we relied on self-reported depression and anxiety symptoms. There was no verification of symptoms by a clinician. While this is an expansion of our prior work that relied on self-disclosure of diagnoses, there are still problems with self-report (Rutter et al., 2023), as compared to clinician-based severity ratings. Clinical diagnoses of GAD and depression were not confirmed in our sample. Our measure of GAD was from the emerging measures list for *DSM-5*, and is not the gold standard for GAD like the GAD-7. Second, our sample was drawn from an online Qualtrics panel, and data may be limited by low-incentive to complete the research accurately (Douglas et al., 2023). Third, we had low proportions of individuals with extreme levels of depression and anxiety symptoms. While this is to be expected, the bootstrapping produced large confidence intervals, which limited interpretation of CDS at extreme levels of symptoms. A final limitation is that our CDS lexicon has not been tested against human coders of the Twitter messages. Human annotation and extension with additional examples constructed by LLMs is a next logical step of this work.

This work has implications for the future of evidence-based treatments including online interventions that engage individuals with internalizing symptoms, but much more work is needed to determine the ways that NLP/AI can meaningfully impact treatment. Characterizing the relationship between distorted language and depression and anxiety symptoms may help in the development of automated interventions such as chatbots, see (Ahmed et al., 2023) for a recent review. Moreover, the extent to which CDS prevalence occurs in the population at large can be used as a passive index of vulnerability to depression and anxiety disorders. This vulnerability index is expected to change as individuals progress through treatment, or go through major stressors, such as a global pandemic, but this is yet to be explored. Using online language to distinguish trajectories of symptoms in individuals with internalizing symptoms is still a relatively novel research area (Stade et al., 2023, 2024). More work is needed to understand cognitive distortions in colloquial language use (Bollen et al., 2021) and how these distortions can change with age, symptoms, context, and treatment.

Author Contributions L.R. and A.E. wrote the main manuscript text and A. E. prepared Figs. 1, 2 and 3. M.t.T. edited figures and the corresponding manuscript text. All authors reviewed the manuscript.

Funding L. Rutter is supported by a K01 award from the National Institute of Aging [K01AG086598]. This work is also supported by the National Institute of Mental Health [R01MH135502].

Data Availability No datasets were generated or analysed during the current study.

Declarations

Competing Interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Ahmed, A., Hassan, A., Aziz, S., Abd-Alrazaq, A. A., Ali, N., Alzubaidi, M., Al-Thani, D., Elhusein, B., Siddig, M. A., Ahmed, M., et al. (2023). Chatbot features for anxiety and depression: A scoping review. *Health Informatics Journal*, 29(1), 14604582221146719.
- Al-Mosaiwi, M., & Johnstone, T. (2018). In an absolute state: Elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. *Clinical Psychological Science*, 6(4), 529–542.
- Babu, N. V., & Kanaga, E. G. M. (2022). Sentiment analysis in social media data for depression detection using artificial intelligence: A review. *SN Computer Science*, 3, 1–20.
- Bathina, K. C., Thij, T., Lorenzo-Luaces, M., Rutter, L., L. A., & Bollen, J. (2021). Individuals with depression express more distorted thinking on social media. *Nature Human Behaviour*, 5(4), 458–466.
- Beck, A. T. (1963). Thinking and depression: I. idiosyncratic content and cognitive distortions. *Archives of General Psychiatry*, 9(4), 324–333.
- Beck, A. T., & Clark, D. A. (1988). Anxiety and depression: An information processing perspective. *Anxiety Research*, 1(1), 23–36. <https://doi.org/10.1080/10615808808248218>
- Berkman, L. F., Berkman, C. S., Kasl, S., Freeman Jr, D. H., Leo, L., Ostfeld, A. M., Cornoni-Huntley, J., & Brody, J. A. (1986). Depressive symptoms in relation to physical health and functioning in the elderly. *American Journal of Epidemiology*, 124(3), 372–388.
- Bernard, J. D., Baddeley, J. L., Rodriguez, B. F., & Burke, P. A. (2016). Depression, language, and affect: An examination of the influence

- of baseline depression and affect induction on language. *Journal of Language and Social Psychology*, 35(3), 317–326.
- Black, D., Riskind, J. H., & Kleiman, E. M. (2010). Lifetime history of anxiety and mood disorders predicted by cognitive vulnerability to anxiety. *International Journal of Cognitive Therapy*, 3(3), 215–227.
- Bollen, J., Ten Thij, M., Breithaupt, F., Barron, A. T., Rutter, L. A., Lorenzo-Luaces, L., & Scheffer, M. (2021). Historical language records reveal a surge of cognitive distortions in recent decades. *Proceedings of the National Academy of Sciences*, 118(30), e2102061118.
- Deshpande, M., & Rao, V. (2017). Depression detection using emotion artificial intelligence. *2017 international conference on intelligent sustainable systems (iciss)*, 858–862.
- DeSouza, D. D., Robin, J., Gumus, M., & Yeung, A. (2021). Natural language processing as an emerging tool to detect late-life depression. *Frontiers in Psychiatry*, 12, 719125.
- Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between mturk, prolific, cloudresearch, qualtrics, and sona. *Plos One*, 18(3), e0279720.
- Duffy, A., Horrocks, J., Doucette, S., Keown-Stoneman, C., McCloskey, S., & Grof, P. (2013). Childhood anxiety: An early predictor of mood disorders in offspring of bipolar parents. *Journal of Affective Disorders*, 150(2), 363–369.
- Edinger, A., Valdez, D., Walsh-Buhi, E., Trueblood, J. S., Lorenzo-Luaces, L., Rutter, L. A., & Bollen, J. (2023). Misinformation and public health messaging in the early stages of the mpox outbreak: Mapping the twitter narrative with deep learning. *Journal of Medical Internet Research*, 25, e43841.
- Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion (Washington D C)*, 7(2), 336–353. <https://doi.org/10.1037/1528-3542.7.2.336>
- Fried, E. I., Coomans, F., & Lorenzo-Luaces, L. (2020). The 341 737 ways of qualifying for the melancholic specifier. *The Lancet Psychiatry*, 7(6), 479–480.
- Granger, D. A., Weisz, J. R., & Kauneckis, D. (1994). Neuroendocrine reactivity, internalizing behavior problems, and control-related cognitions in clinic-referred children and adolescents. *Journal of Abnormal Psychology*, 103(2), 267.
- Greenberg, P. E., Sisitsky, T., Kessler, R. C., Finkelstein, S. N., Berndt, E. R., Davidson, J. R., Ballenger, J. C., & Fyer, A. J. (1999). The economic burden of anxiety disorders in the 1990s. *Journal of Clinical Psychiatry*, 60(7), 427–435.
- Greenberg, P. E., Fournier, A. A., Sisitsky, T., Pike, C. T., & Kessler, R. C. (2015). The economic burden of adults with major depressive disorder in the United States (2005 and 2010). *The Journal of Clinical Psychiatry*, 76(2), 155–162.
- Holtzman, N. S., et al. (2017). A meta-analysis of correlations between depression and first person singular pronoun use. *Journal of Research in Personality*, 68, 63–68.
- Hong, R. Y., & Cheung, M. W. L. (2015). The structure of cognitive vulnerabilities to depression and anxiety: Evidence for a common core etiologic process based on a meta-analytic review. *Clinical Psychological Science*, 3(6), 892–912.
- Hoza, B., Mrug, S., Gerdes, A. C., Hinshaw, S. P., Bukowski, W. M., Gold, J. A., Kraemer, H. C., Pelham Jr, W. E., Wigal, T., & Arnold, L. E. (2005). What aspects of peer relationships are impaired in children with attention-deficit/hyperactivity disorder? *Journal of Consulting and Clinical Psychology*, 73(3), 411.
- Ka'zmierzczak, I., Jakubowska, A., Pietraszkiewicz, A., Zajenowska, A., Lacko, D., Wawer, A., & Sarzyn'ska-Wawer, J. (2024). Natural language sentiment as an indicator of depression and anxiety symptoms: A longitudinal mixed methods study1. *Cognition and Emotion*, 1–10.
- Kelley, S. W., Fisher, A. J., Lee, C. T., Gallagher, E., Hanlon, A. K., Robertson, I. H., & Gillan, C. M. (2023). Elevated emotion network connectivity is associated with fluctuations in depression. *Proceedings of the National Academy of Sciences*, 120(45), e2216499120.
- Konnopka, A., & K'onig, H. (2020). Economic burden of anxiety disorders: A systematic review and meta-analysis. *Pharmacoeconomics*, 38, 25–37.
- Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The phq-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9), 606–613.
- Lawrence, H. R., Schneider, R. A., Rubin, S. B., Matari'c, M. J., McDuff, D. J., & Bell, M. J. (2024). The opportunities and risks of large language models in mental health. *JMIR Mental Health*, 11(1), e59479.
- Le Glaz, A., Haralambous, Y., Kim-Dufor, D. H., Lenca, P., Billot, R., Ryan, T. C., Marsh, J., Devylder, J., Walter, M., Berrouiguet, S., et al. (2021). Machine learning and natural language processing in mental health: Systematic review. *Journal of Medical Internet Research*, 23(5), e15708.
- Lee, M. T., & Theokary, C. (2021). The superstar social media influencer: Exploiting linguistic style and emotional contagion over content? *Journal of Business Research*, 132, 860–871.
- Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*, 79, 328–348. <https://doi.org/10.1016/j.jesp.2018.08.009>
- Lorenzo-Luaces, L. (2015). Heterogeneity in the prognosis of major depression: From the common cold to a highly debilitating and recurrent illness. *Epidemiology and Psychiatric Sciences*, 24(6), 466–472.
- Lorenzo-Luaces, L., German, R. E., & DeRubeis, R. J. (2015). It's complicated: The relation between cognitive change procedures, cognitive change, and symptom change in cognitive therapy for depression. *Clinical Psychology Review*, 41, 3–15.
- Lorenzo-Luaces, L., Keefe, J. R., & DeRubeis, R. J. (2016). Cognitive-behavioral therapy: Nature and relation to non-cognitive behavioral therapy. *Behavior Therapy*, 47(6), 785–803.
- Lorenzo-Luaces, L., Rutter, L. A., & Scalco, M. D. (2020). Carving depression at its joints? Psychometric properties of the sydney melancholia prototype index. *Psychiatry Research*, 293, 113410.
- Lorenzo-Luaces, L., Buss, J. F., & Fried, E. I. (2021). Heterogeneity in major depression and its melancholic and atypical specifiers: A secondary analysis of star* d. *Bmc Psychiatry*, 21, 1–11.
- Lorenzo-Luaces, L., Howard, J., Edinger, A., Yan, H. Y., Rutter, L. A., Valdez, D., Bollen, J., et al. (2022). Sociodemographics and transdiagnostic mental health symptoms in social (studies of online cohorts for internalizing symptoms and language) i and ii: Cross-sectional survey and botometer analysis. *JMIR Formative Research*, 6(10), e39324.
- Milintsevich, K., Sirts, K., & Dias, G. (2023). Towards automatic text-based estimation of depression through symptom prediction. *Brain Informatics*, 10(1), 4.
- Rude, S., Gortner, E. M., & Pennebaker, J. (2004). Language use of depressed and depression-vulnerable college students. *Cognition & Emotion*, 18(8), 1121–1133.
- Rutter, L. A., & Brown, T. A. (2015). Reliability and validity of the dimensional features of generalized anxiety disorder. *Journal of Anxiety Disorders*, 29, 1–6.
- Rutter, L. A., Scheuer, L., Vahia, I. V., Forester, B. P., Smoller, J. W., & Germaine, L. (2019). Emotion sensitivity and self-reported symptoms of generalized anxiety disorder across the lifespan: A population-based sample approach. *Brain and Behavior*, 9(6), e01282.
- Rutter, L. A., Norton, D. J., & Brown, T. A. (2020a). The impact of self-reported depression severity and age on facial emotion

- recognition in outpatients with anxiety and mood disorders. *Journal of Psychopathology and Behavioral Assessment*, 42(1), 86–92.
- Rutter, L. A., Passell, E., Scheuer, L., & Germine, L. (2020b). Depression severity is associated with impaired facial emotion processing in a large international sample. *Journal of Affective Disorders*, 275, 175–179.
- Rutter, L. A., Thompson, H. M., Howard, J., Riley, T. N., De Jesu's-Romero, R., & Lorenzo-Luaces, L. (2021). Social media use, physical activity, and internalizing symptoms in adolescence: Cross-sectional analysis. *JMIR Mental Health*, 8(9), e26134.
- Rutter, L. A., Lind, C., Howard, J., Lakhan, P., & Germine, L. (2022). Posttraumatic stress symptom severity is associated with impaired processing of emotional faces in a large international sample. *Journal of Traumatic Stress*, 35(4), 1263–1272.
- Rutter, L. A., Howard, J., Lakhan, P., Valdez, D., Bollen, J., & Lorenzo-Luaces, L. (2023). I haven't been diagnosed, but i should be—insight into self-diagnoses of common mental health disorders: Cross-sectional study. *JMIR Formative Research*, 7(1), e39206.
- Rutter, L. A., Thij, T., Lorenzo-Luaces, M., Valdez, L., D., & Bollen, J. (2024). Negative affect variability differs between anxiety and depression on social media. *Plos One*, 19(2), e0272107.
- Sayyadiharikandeh, S., Varol, O., Yang, K-C., Flammini, A., & Menczer, F. (2020). Detection of Novel Social Bots by Ensembles of Specialized Classifiers. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management (CIKM '20). Association for Computing Machinery, New York, NY, USA, 2725–2732. <https://doi.org/10.1145/3340531.3412698>
- Stade, E. C., Ungar, L., Eichstaedt, J. C., Sherman, G., & Ruscio, A. M. (2023). Depression and anxiety have distinct and overlapping language patterns: Results from a clinical interview. *Journal of psychopathology and clinical science*.
- Stade, E. C., Stirman, S. W., Ungar, L. H., Boland, C. L., Schwartz, H. A., Yaden, D. B., Sedoc, J., DeRubeis, R. J., Willer, R., & Eichstaedt, J. C. (2024). Large language models could change the future of behavioral healthcare: A proposal for responsible development and evaluation. *NPJ Mental Health Research*, 3(1), 12.
- Tackman, A. M., Sbarra, D. A., Carey, A. L., Donnellan, M. B., Horn, A. B., Holtzman, N. S., Edwards, T. S., Pennebaker, J. W., & Mehl, M. R. (2019). Depression, negative emotionality, and self-referential language: A multi-lab, multi-measure, and multi-language-task research synthesis. *Journal of Personality and Social Psychology*, 116(5), 817.
- Ten Thij, M., Bathina, K., Rutter, L. A., Lorenzo-Luaces, L., van de Leemput, I. A., Scheffer, M., & Bollen, J. (2020). Depression alters the circadian pattern of online activity. *Scientific Reports*, 10(1), 1–10.
- Tong, A., Du, D. Z., & Wu, W. (2018). On misinformation containment in online social networks. *Advances in neural information processing systems*, 31.
- Young, E., & Korszun, A. (2010). Sex, trauma, stress hormones and depression. *Molecular Psychiatry*, 15(1), 23–28.

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