

Original Paper

Emotional Expression on Social Media Support Forums for Substance Cessation: Observational Study of Text-Based Reddit Posts

Genevieve Yang^{1,2}, MD, PhD; Sarah G King², BA; Hung-Mo Lin^{3,4}, SCD; Rita Z Goldstein^{1,2}, PhD

¹Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York City, NY, United States

²Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York City, NY, United States

³Department of Anesthesiology, Yale School of Medicine, Yale University, New Haven, CT, United States

⁴Yale Center for Analytical Sciences, Yale School of Public Health, Yale University, New Haven, CT, United States

Corresponding Author:

Rita Z Goldstein, PhD

Department of Psychiatry

Icahn School of Medicine at Mount Sinai

1 Gustave L Levy Pl

New York City, NY, 10029

United States

Phone: 1 212 659 8886

Email: rita.goldstein@mssm.edu

Abstract

Background: Substance use disorder is characterized by distinct cognitive processes involved in emotion regulation as well as unique emotional experiences related to the relapsing cycle of drug use and recovery. Web-based communities and the posts they generate represent an unprecedented resource for studying subjective emotional experiences, capturing population types and sizes not typically available in the laboratory. Here, we mined text data from Reddit, a social media website that hosts discussions from pseudonymous users on specific topic forums, including forums for individuals who are trying to abstain from using drugs, to explore the putative specificity of the emotional experience of substance cessation.

Objective: An important motivation for this study was to investigate transdiagnostic clues that could ultimately be used for mental health outreach. Specifically, we aimed to characterize the emotions associated with cessation of 3 major substances and compare them to emotional experiences reported in nonsubstance cessation posts, including on forums related to psychiatric conditions of high comorbidity with addiction.

Methods: Raw text from 2 million posts made, respectively, in the fall of 2020 (discovery data set) and fall of 2019 (replication data set) were obtained from 394 forums hosted by Reddit through the application programming interface. We quantified emotion word frequencies in 3 substance cessation forums for alcohol, nicotine, and cannabis topic categories and performed comparisons with general forums. Emotion word frequencies were classified into distinct categories and represented as a multidimensional emotion vector for each forum. We further quantified the degree of emotional resemblance between different forums by computing cosine similarity on these vectorized representations. For substance cessation posts with self-reported time since last use, we explored changes in the use of emotion words as a function of abstinence duration.

Results: Compared to posts from general forums, substance cessation posts showed more expressions of anxiety, disgust, pride, and gratitude words. “Anxiety” emotion words were attenuated for abstinence durations >100 days compared to shorter durations ($t_{12}=3.08$, 2-tailed; $P=.001$). The cosine similarity analysis identified an emotion profile preferentially expressed in the cessation posts across substances, with lesser but still prominent similarities to posts about social anxiety and attention-deficit/hyperactivity disorder. These results were replicated in the 2019 (pre-COVID-19) data and were distinct from control analyses using nonemotion words.

Conclusions: We identified a unique subjective experience phenotype of emotions associated with the cessation of 3 major substances, replicable across 2 time periods, with changes as a function of abstinence duration. Although to a lesser extent, this phenotype also quantifiably resembled the emotion phenomenology of other relevant subjective experiences (social anxiety and

attention-deficit/hyperactivity disorder). Taken together, these transdiagnostic results suggest a novel approach for the future identification of at-risk populations, allowing for the development and deployment of specific and timely interventions.

(*J Med Internet Res* 2023;25:e45267) doi: [10.2196/45267](https://doi.org/10.2196/45267)

KEYWORDS

sentiment analysis; text mining; addiction phenotype; subjective experience phenotype; naturalistic big data; natural language processing; phenomenology; experience sampling

Introduction

Background

Substance use disorders afflict millions of Americans, exacting an immeasurable toll on individuals and communities. A core feature of drug addiction is the emergence of withdrawal symptoms characterized by negative affective states that, through negative reinforcement, contribute to a relapsing cycle of continued drug use despite adverse consequences [1,2]. Emerging evidence suggests that subjective emotional experiences while using a drug [3,4], and during abstinence [5,6], may be a powerful predictor of problem drug use behavior both for that drug and, notably, other substances as well [7], underscoring the significance of emotion regulation and expression to substance use disorders in general. An important challenge in the characterization and treatment of drug addiction is to identify the role-specific emotions play in the escalation of substance use and in its cessation.

Understanding the subjective experiences of substance use and cessation might also provide valuable insights into the epidemiological links between addiction and related disorders. When compared with the general population, problem drug use and addiction are highly prevalent in clinical populations with nonsubstance-related mental health disorders that encompass impairments in emotion regulation, especially mood disorders (including depression), anxiety, psychotic disorders, and attention-deficit/hyperactivity disorder (ADHD) [8]. The mechanisms underlying such comorbidities with substance use disorders are poorly understood, although they likely reflect clusters of shared neurobiological predispositions and environmental risk factors [9,10]. Individuals with comorbid substance and nonsubstance disorders frequently experience poorer abstinence-based treatment outcomes [8], highlighting the therapeutic importance of transdiagnostic approaches to studying drug addiction.

Prior Work

A key element of subjective experiences is emotional content, which can be extracted from verbal reports such as in speech or text. Importantly, social media has become a ubiquitous channel for text-based public discourse on topics including

mental illness and drug use or addiction. A rapidly increasing number of studies are now tapping into publicly accessible, user-generated, and user-friendly social media content and databases, which facilitate the gathering of large amounts of targeted data, capturing population types and sample sizes that are not typically amenable to laboratory-based studies. Such studies have been applied to understanding current and epidemiological trends in public health [11], such as attitudes toward COVID-19 [12,13] and substance use disorders [14], exposing patterns of human behavior that may not be otherwise captured through standard experimental and clinical tools. As a result, web-based discussion forums such as the social media platform Reddit have become a vital resource for the large-scale collection of ever-expanding naturalistic, and topical, health data.

Such data sets are highly suitable for natural language processing (NLP) analysis methods, providing ubiquitous and readily accessible measures of human behavior that can be mined for public opinion and marketing purposes, but also public health research. On Reddit, pseudonymous users submit posts and post replies to single-topic discussion forums called subreddits (see [Table 1](#) for an example post), including a variety of mental health subreddits, with some that offer support for substance use cessation. Using a variety of NLP-based analyses, studies using data from Reddit (as well as Twitter, Facebook, and Instagram) have effectively applied a naturalistic lens to how individuals communicate about drugs, including their own consumption patterns and attempts to quit or abstain [15-17]. A potentially powerful application of these studies is the ability to predict future drug use, both at the population (ie, epidemiological trends in specific drug classes) and individual (ie, the propensity to relapse) levels. Considering the already established role that subjective emotional experiences play in addiction vulnerability and problem drug use, the emotional content of spontaneously generated self-reports might provide an especially valuable predictor of an individual's future drug use and markers of recovery, which have been relatively neglected in the current addiction literature. Importantly, anonymous social media data can capture users' candid experiences with substances and substance use or cessation in real time, including individual emotional accounts associated with success or failure to abstain.

Table 1. Example posts from substance cessation subreddits.

Subreddit	Title	Text
r/stopdrinking	I just realized that Thanksgiving will be exactly 60 days sober for me	A year ago I couldn't even imagine a life free from alcohol. I have so much to be thankful for even with all the nonsense going on these days.
r/stopsmoking	Here we go again	My fourth time seriously trying to quit! I'm at 4 days 8 hours right now and I am very proud of myself. It hasn't been easy though. Anyone else have experience with...
r/leaves	5 months 23 days	I never thought I'd make it this far but I'm doing great. I'm feeling in control, I don't need to smoke any time I get bored or sad or lonely or stressed, I have money to spend...

Objective

Using 2 million Reddit posts each from the fall of 2020 and the fall of 2019 across 394 topic categories, we studied emotional content in posts from 3 major substance cessation forums (alcohol, nicotine, and cannabis) as compared to nonsubstance cessation posts, with the following objectives: (1) to identify unique patterns and specificity by comparing results to the largest nonsubstance cessation subreddits (those exceeding 1 million members), allowing for transdiagnostic comparisons; (2) to identify general patterns by exploring cross-substance similarities in emotions expressed in the 3 substance cessation forums; (3) to characterize changes in patterns of emotion expression as a function of self-reported abstinence; and (4) to examine the reproducibility of findings by comparing 2 data sets (a post-COVID-19 discovery data set and a pre-COVID-19 replication data set).

Methods

Feature Selection

We selected emotion words of interest a priori and independent of the Reddit data. To ensure reasonable coverage of common emotion categories, we used multiple resources, including the primary and secondary emotions proposed by Plutchik [18]. In total, 21 emotion categories were selected for the curated emotion word bank (Table S1 in [Multimedia Appendix 1](#) [18-23]), and 15 categories for a control analysis using the curated time word bank (Table S2 in [Multimedia Appendix 1](#)). Using the *parsing.preprocessing* package from the Gensim open-source Python library created by Řehůřek and Sojka [24], raw text from Reddit posts were preprocessed to convert all letters to lowercase and strip all punctuation, multiple white spaces, and HTML tags (eg, tags for bolding or italicization).

Search Strategy and Selection Criteria

The exemplar substance cessation subreddits for the 3 most commonly used drug classes were identified by using internet searches for “Reddit,” “addiction,” and the substance of interest, then choosing the subreddit result with the largest membership base. The r/stopdrinking subreddit had 256,000 members, r/stopsmoking 113,000 members, and r/leaves 153,000 members as of November 2020. In addition, we included a total of 391 comparison subreddits, each with greater than 1 million members, to capture a subject pool that might resemble the general population for a control sample, with the 3 selected substance cessation subreddits. Thus, the discovery data set comprised a total of 394 subreddits and nearly 2 million unique posts. For each subreddit, the pushshift.io Reddit application programming interface was used to extract 5025 consecutive

posts (see [Table 1](#) for an example) from late November 2020 through early January 2021.

Subreddits in which fewer than 3000 of the 5025 posts contained body text were excluded. Implementing this arbitrary cutoff (representing a decision to use as many viable posts as possible), 105 (of 394) subreddits remained (in addition to the 3 exemplar substance cessation subreddits). The subreddit with the fewest text posts passing this cutoff contained 3003 text posts (out of the original 5025). Therefore, to maintain a consistent number of posts across subreddits, only the most recent 3003 textual posts from each sample were used for analysis. We estimated that approximately 2400 unique users per subreddit generated these posts (see “Estimation of Number of Unique Subjects” in [Multimedia Appendix 1](#) for details). The replication data set of pre-COVID-19 posts included a starting sample of the same 105 comparison subreddits from November 2019, which underwent identical quality control for textual content, resulting in a sample of 92 (13 subreddits were excluded due to low text content; Figure S1 in [Multimedia Appendix 1](#)).

For each subreddit's 3003 posts, we quantified the number of emotion word bank matches to produce an emotion score (Figure S2 in [Multimedia Appendix 1](#)). Using a median split, we then excluded emotion-poor subreddits (those below the median). All substance cessation support subreddits in the study met the criteria for a “high emotion” classification (above the median split) when evaluated together with the 105 comparison subreddits, forming a final sample of 54 subreddits for emotion analysis (47 in the replication sample; Figure S1 in [Multimedia Appendix 1](#)).

To assess the potential effects of abstinence duration, we extracted an abstinence-tagged data set including a subset of posts from r/stopdrinking and r/stopsmoking, which allow users to self-report the number of days since their last substance use (the r/leaves subreddit does not offer this feature). First, the most recent 15,025 consecutive posts as of late November 2020 were extracted from each of these 2 subreddits and checked for the presence of metadata for abstinence duration at the time of the post. After excluding image-only posts (as described above) and posts lacking abstinence metadata, 6595 posts remained for r/stopdrinking and 1205 for r/stopsmoking. To maintain the same number of posts across subreddits, only the most recent 1205 posts from the sample of 6595 were used for r/stopdrinking.

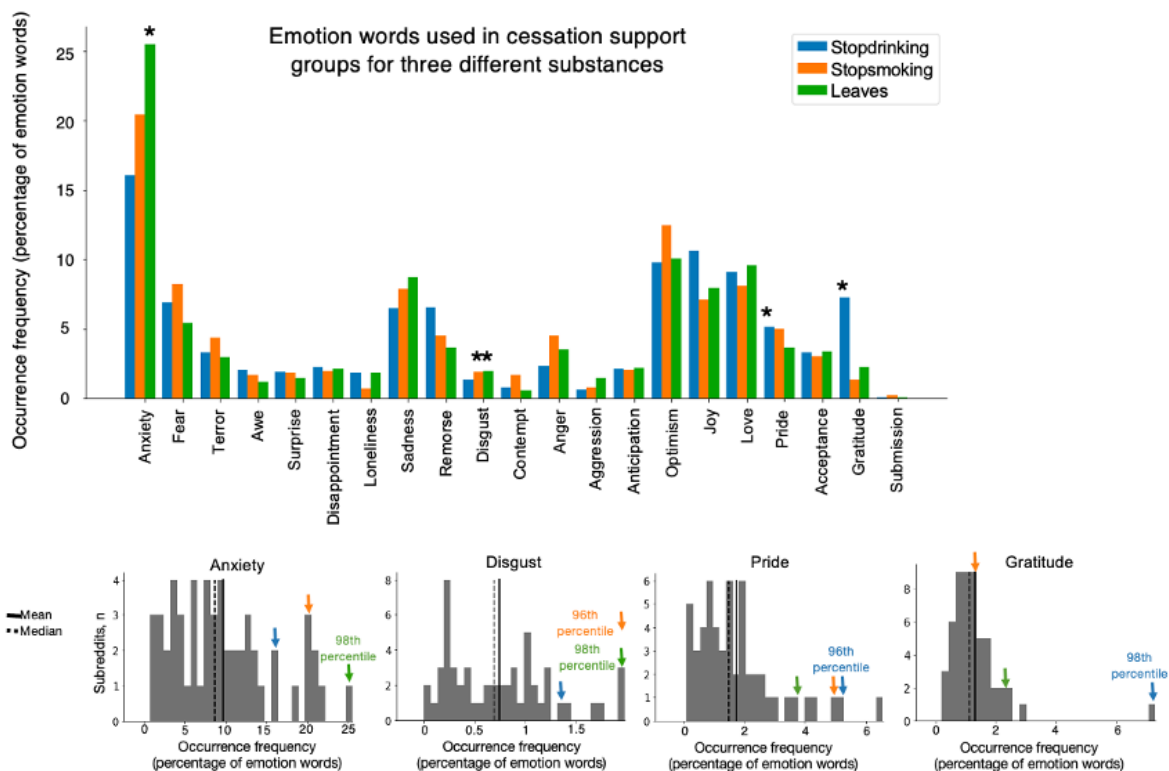
Choice of Primary Measure and Statistical Approach

Our primary measures were occurrence frequency and cosine similarity. For the emotion analysis, occurrence frequency was a count of the number of times words appear in each emotion

category, normalized by the total number of emotion word matches, producing a percentage value for each of the 21 emotion categories (see “Emotion Word Bank” in [Multimedia Appendix 1](#) for details). This 21-item list of percentages was computed for each subreddit independently of other subreddits. Subreddits were considered outliers on a given emotion if their word frequency fell in the 95th percentile on that emotion compared to the control sample of large subreddits of at least 1 million members. We similarly used occurrence frequency and time word categories for the time word bank. To quantify the similarity between a given pair of subreddits, we transformed the 21-item lists of emotion occurrence frequencies into 21-dimensional vectors for each subreddit and computed the cosine of the angle (cosine similarity) between each pair of vectors, corresponding to a pair of subreddits (see “Cosine Similarity” in [Multimedia Appendix 1](#) for details). Similar procedures were used for the time word vectors (see “Time Word Bank” in [Multimedia Appendix 1](#) for details).

To examine the effects of emotion (using all 21 categories) on word occurrence frequency as a function of abstinence duration, we used a 2-way ANOVA incorporating 2 independent categorical variables: emotion and abstinence duration. Due to limited post data for longer abstinence durations, we applied a cutoff of 1000 days for this analysis, with posts consolidated into 15 time bins. The abstinence factor was then divided into 2 levels denoting short-term (<100 days, 9 time bins) and long-term (>100 days, 6 time bins) abstinence. Significant interactions between emotion and abstinence duration were followed up with post hoc *t* tests. To limit multiple comparisons, we restricted the post hoc 2-sided independent samples *t* tests (comparing posts tagged with self-reported long-term abstinence vs posts tagged with short-term abstinence) to the 4 emotions identified in [Figure 1](#), with Bonferroni correction ($\alpha=.05/4$ or $P<.01$).

Figure 1. Emotion word analysis of posts from cessation support subreddits and control subreddits. The substance cessation support subreddits were compared with high-membership general interest subreddits with emotion-rich text (see Figure S1 in [Multimedia Appendix 1](#)). Top plot represents occurrence frequency of each emotion word category. Histograms show the 4 emotion words on which the substance cessation subreddits demonstrated outlier properties. Colored arrows indicate locations of the substance cessation subreddits within the larger distribution of high-emotion subreddits. Blue = r/stopdrinking, orange = r/stopsmoking, green = r/leaves. *denotes that the subreddit was an outlier (>95th percentile) in use of the target emotion words compared to control general topic subreddits.



Ethical Considerations

Users of Reddit’s web-based forums are made aware that their posts are publicly accessible in the website’s terms and conditions. All data were collected through the free and open-access application processing interface, which does not provide personal identifying information such as legal names, locations, or IP addresses. For the purposes of this study, we did not participate in discussions on the forums, and there was

therefore no ethical consideration to inform users that their posts may be used for research. For these reasons, this study received an exemption status from Mount Sinai’s Institutional Review Board.

Results

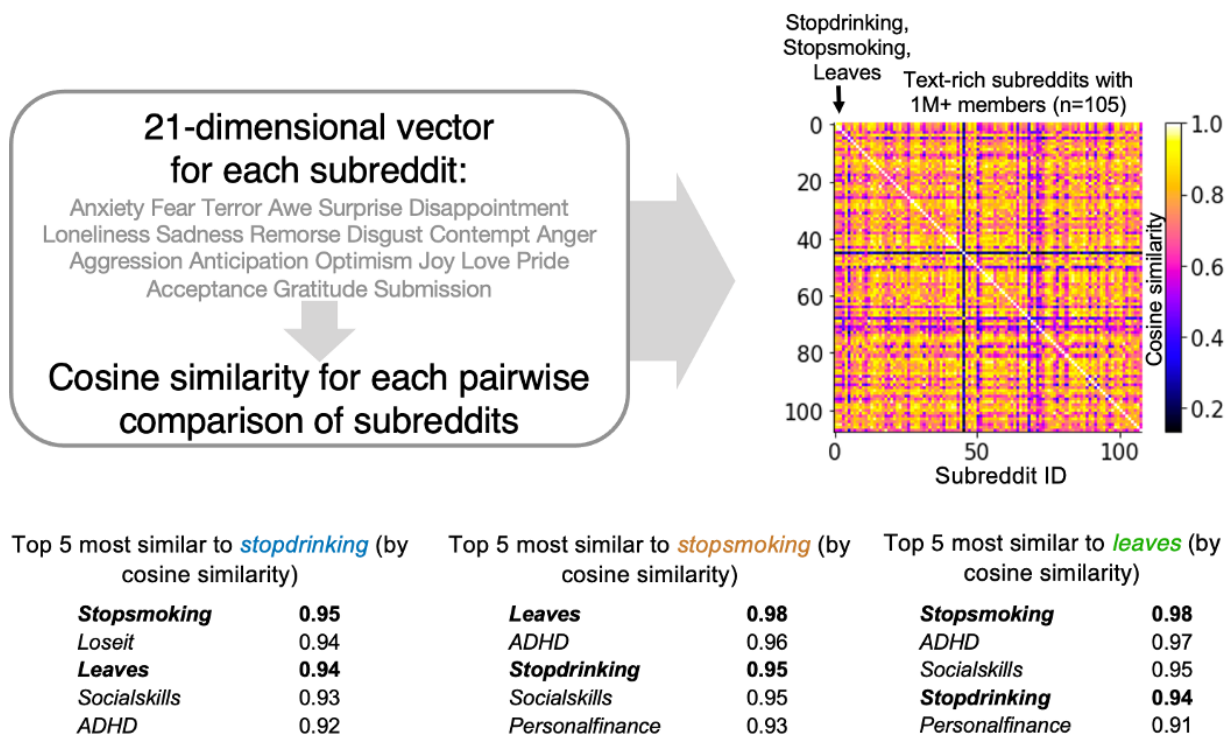
Emotional Content in Substance Cessation Subreddits

The emotion compositions of the 3 substance use cessation subreddits (r/stopdrinking, r/stopsmoking, and r/leaves) similarly scored >15% (top 8/54) for “anxiety” and around 10% (top 5/54) for “sadness,” “optimism,” “joy,” and “love” (Figure 1). Furthermore, on 4 of the emotion categories, “anxiety,” “disgust,” “pride,” and “gratitude,” the substance cessation subreddits showed outlier levels of word occurrence frequency (in the 95th percentile, 3/54) compared to the full sample of subreddits (Figure 1). Specifically, the nicotine cessation subreddit (r/stopsmoking) showed high “disgust,” the cannabis cessation subreddit (r/leaves) showed a high representation for both “disgust” and “anxiety,” and the alcohol cessation subreddit (r/stopdrinking) showed a high representation for “pride” and

“gratitude.” Although r/stopsmoking was also high in the “gratitude” emotion, this trend did not exceed the 95th percentile.

Cosine similarity analyses were performed by comparing the emotion profiles for every possible subreddit pair in our data set (Figure 2). Notably, the substance use cessation subreddits appeared consistently in the list of top 5 most emotionally similar subreddits for each of the other substance use cessation subreddits (Figure 2). Indeed, for each of these subreddits, the next most emotionally similar subreddit was one of the other substance use cessation subreddits. Additional subreddits appearing in the top 5 lists for all 3 substance cessation subreddits were r/ADHD and r/socialskills. Additionally, r/loseit, a weight-loss support subreddit, appeared in the top 5 for r/stopdrinking, and r/personalfinance, a subreddit for financial advice, appeared for both r/stopsmoking and r/leaves.

Figure 2. Emotional similarity of substance cessation and control subreddits. Emotion word occurrence frequencies were computed and expressed as a 21-dimensional emotion vector for each subreddit. Gray words list the emotion dimensions. Cosine similarity was computed between each pair of subreddits with respect to their emotion vectors. Heatmap illustrates the cosine similarity for pair-wise comparison analysis among all subreddits. The exemplar subreddits are at the top left (in order: r/stopdrinking, then r/stopsmoking, then r/leaves), followed by the 105 comparison subreddits in order of decreasing member size, starting with the largest: r/Showerthoughts, with over 22 million members (note: the final analysis included only the 54 most emotion dense subreddits). For each of the substance cessation subreddits, the top 5 most emotionally similar subreddits are shown with accompanying cosine similarity scores (bottom). ADHD: attention-deficit/hyperactivity disorder.



Control Analyses: Time Words in Substance Cessation Subreddits

Cosine similarity analyses were performed by comparing the time word profiles for every possible subreddit pair in our data set similar to the emotion word analysis. Notably, the emotional similarity between the substance use cessation subreddits did not replicate for time words. Specifically, the most similar subreddits to both r/stopdrinking and r/leaves were not substance cessation subreddits (r/leaves remained the most similar to r/stopdrinking), while r/stopdrinking and r/leaves appeared in

each other’s list of top 5 most similar subreddits, r/stopsmoking did not appear in either of the other 2 subreddits’ lists or vice versa (see Figure S2 in Multimedia Appendix 1). Interestingly, in these time word analyses, only 1 substance use cessation subreddit displayed outlier properties: r/stopsmoking for “morning” words (see Figure S3 in Multimedia Appendix 1).

Emotional Content and Abstinence

Posting patterns were assessed for subreddits with self-reported abstinence duration metadata (r/stopdrinking and r/stopsmoking; Figure 3). Both subreddits showed a disproportionately high

post volume in the first 24 hours of abstinence. That is, in both subreddits, at least 7% (462/1205) of all posts were made within the first 24 hours of abstinence, which is more than 2 orders of magnitude greater than chance expectations (ie, given that the abstinence period range extended to over 10,000 days for each subreddit, if users were equally likely to post at any abstinence duration, only 0.01%, $n=1$, of posts would be expected to be made within the first 24 hours). In addition, for both subreddits, posting tendencies increased around the 100th day of abstinence.

Given the high frequency of posting within the first 24 hours of abstinence, we quantified the emotion profiles of posts during this period (Figure 3). For both r/stopdrinking and r/stopsmoking, the 2 most highly represented emotions on the first day of abstinence were “anxiety” and “optimism.” The next 3 most highly represented emotions were different for the 2 subreddits: “remorse,” “love,” and “sadness” for r/stopdrinking and “terror,” “disgust,” and “fear” for r/stopsmoking.

Importantly, ANOVA results examining emotion category and abstinence duration relationships for the r/stopdrinking data showed a main effect of emotion category ($F_{20,1}=156.8$; $P<.001$),

no main effect of abstinence duration ($F_{20,1}=0.00$; $P=1.00$), and a significant interaction between emotion category and abstinence duration (interaction effect $F_{20,1}=3.80$; $P<.001$), suggesting a shift in emotion word expression at longer compared to shorter abstinence durations (Figure 4). Post hoc t tests on the r/stopdrinking data restricted to the original 4 emotions of interest identified in Figure 1 (“anxiety,” “disgust,” “pride,” and “gratitude”) confirmed that the “anxiety” emotion was expressed significantly less for abstinence durations greater than 100 days compared to shorter durations ($t_{12}=3.08$, 2-tailed; $P=.01$). The r/stopdrinking data also showed higher “gratitude” emotion expression for longer abstinence durations (greater than 100 days), but this effect did not survive Bonferroni correction for multiple comparisons ($t_{13}=-2.44$, 2-tailed; $P=.03$). The ANOVA results for the r/stopsmoking data showed a main effect of emotion ($F_{20,1}=20.0$; $P<.001$), no main effect of abstinence duration ($F_{20,1}=0.00$; $P=.99$), and no effect of the interaction between emotion and abstinence duration ($F_{20,1}=1.45$; $P=.09$).

Figure 3. Posts on substance cessation subreddits in the first 24 hours of abstinence expressing anxiety and optimism. Analysis of post metadata on self-reported abstinence from 2 substance cessation subreddits, r/stopdrinking and r/stopsmoking. Left plot shows frequencies of posts by length of abstinence. The x-axis is logarithmic and extends to 10,000 days. The table shows word occurrence frequencies in each emotion category restricted to abstinence-tagged posts made within the first 24 hours of abstinence. The top 5 most prevalent emotion word categories are in bold format. Percentage scores express the occurrence frequency of words from the target emotion category relative to the total frequency of any word match for the emotion word bank.

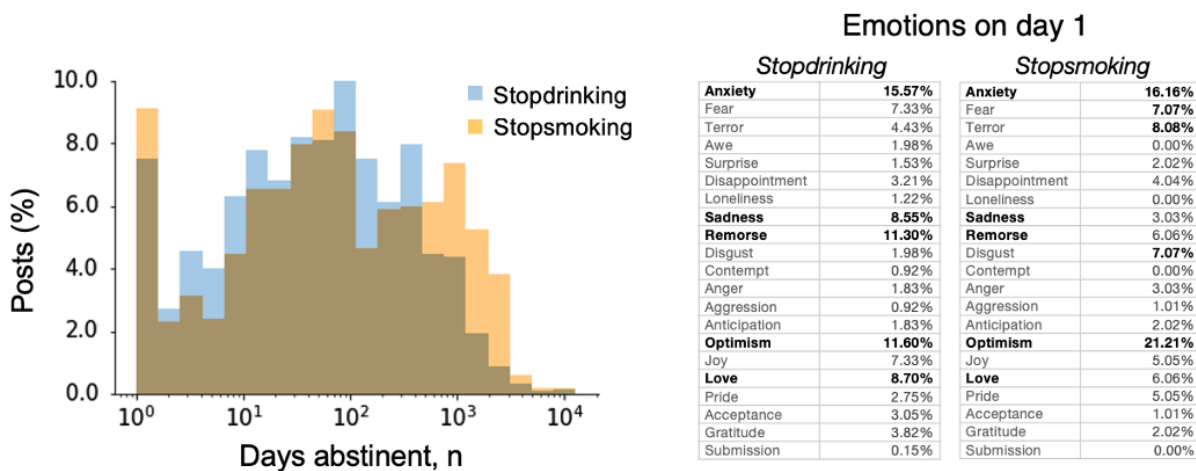


Figure 4. Emotions expressed during the first 1000 days of abstinence. Analysis of post metadata on self-reported abstinence from 2 substance cessation subreddits. Occurrence frequency is shown for emotion words from the 4 emotion categories identified in Figure 1 for short-term (<100 days) and long-term (>100 days) abstinence time points. Long-term abstinence showed a decrease in “anxiety” and a trend toward an increase in “gratitude” emotion expression as compared with the short-term time points for r/stopdrinking. The legend applies to both plots. *significant at $P<.05$, Bonferroni corrected, **approached significance, did not survive Bonferroni correction.

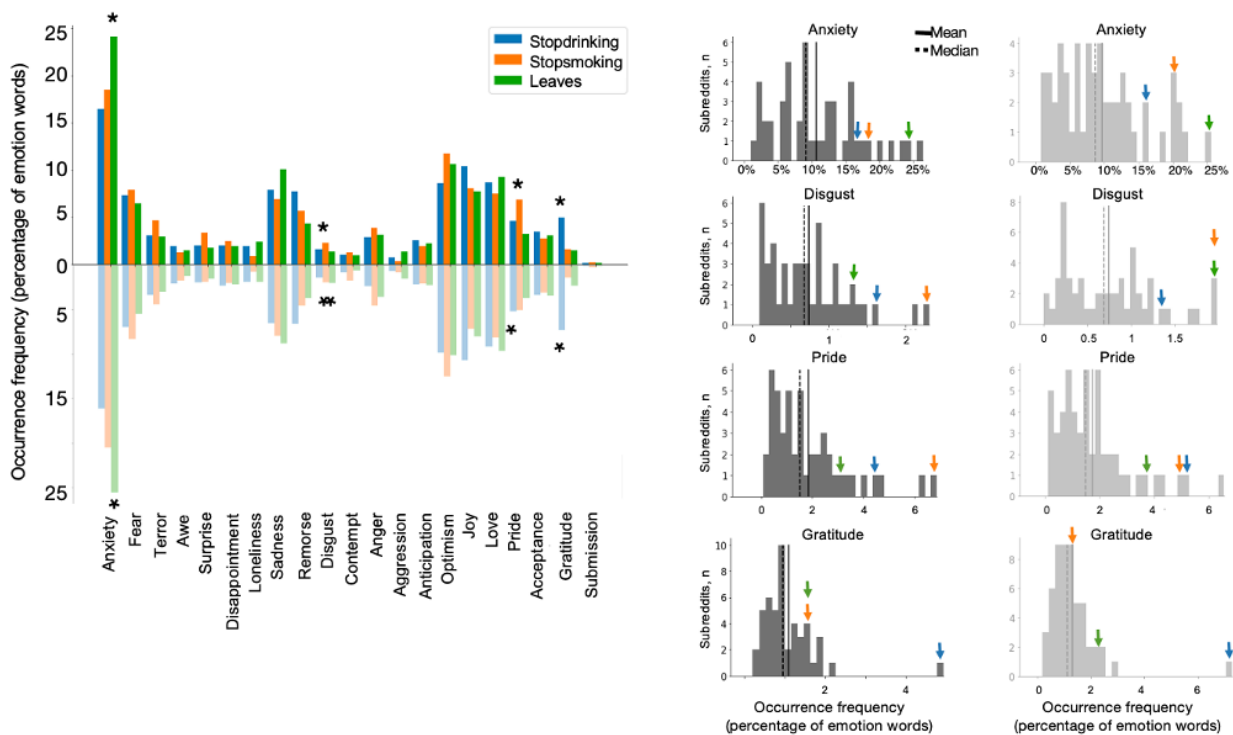


Replication Analysis

Results from the replication data set were compared with the discovery data set (Figure 5). Similar to the previous results, the same 4 outlier emotions were identified in the replication data set. At least one of the substance cessation subreddits showed outlier representation in the top 95th percentile (2/47)

for “anxiety,” “disgust,” “pride,” and “gratitude” emotion categories relative to comparison subreddits. Specifically, the outlier status remained robust for r/leaves on “anxiety,” r/stopsmoking on “disgust,” and r/stopdrinking on “gratitude.” The emotion word cosine similarity showed similar results in this replication data set compared to the discovery data set (see Figure S4 in Multimedia Appendix 1).

Figure 5. Emotion word analysis in the replication data set. Left plot represents emotion word occurrence frequencies for each exemplar substance cessation subreddit, using posts from the replication data set. Occurrence frequencies are expressed as a percentage of all emotion words counted in that subreddit’s posts. The corresponding graph from the discovery data set (Figure 1) is reproduced below the replication data set results as a translucent “reflection” for ease of visual comparison. Histograms show the 4 emotion categories on which the substance cessation subreddits demonstrated outlier properties. Colored arrows indicate locations of the substance cessation subreddits within the larger distribution of high emotion subreddits. Blue = r/stopdrinking, orange = r/stopsmoking, green = r/leaves. *denotes that the subreddit was an outlier (>95th percentile) in use of the target emotion words compared to general topic subreddits.



Discussion

Principal Results

Here, we used spontaneous, self-directed reports provided under relative anonymity in a naturalistic setting (via public forums available on Reddit) generated by hundreds of thousands of subjects to examine the phenomenology of the subjective experience of substance use cessation across 3 of the most common substances of abuse. To apply a transdiagnostic lens, we chose emotion words instead of more addiction-specific features, allowing us to contrast results against the control population of Reddit users. We also compared results as a function of time since last reported drug use and across 2 time periods. Our results suggest an emotional signature preferentially expressed across the alcohol, nicotine, and cannabis cessation subreddits, displaying outlier properties on 4 emotions. These results replicated across 2 time periods (pre- and post-COVID-19), demonstrating continuity in the observed effects over time despite major societal shifts imposed by the pandemic that impacted drug use in general [25]. Significant interactions with time suggested that the identified emotional signature may change with abstinence duration. Although distinct from the control population, this emotion profile showed notable similarity to a few other nonsubstance-themed populations (eg, those plausibly experiencing ADHD in r/ADHD, social anxiety in r/socialskills, financial stress in r/personalfinance, or attempting weight loss in r/loseit), suggestive of transdiagnostic clinical relevance as well as common demographic characteristics.

First, we captured 4 emotions (“disgust,” “anxiety,” “pride,” and “gratitude”) that were expressed within the 3 substance cessation subreddits at a level exceeding 95% (top 3/54) of all included subreddits. The smoking cessation subreddit’s outlier status on “disgust” is consistent with research supporting the role of disgust (as opposed to other emotions, such as health-related anxiety) in reducing craving in smokers [26-29]. Notably, cocaine is another stimulant drug of abuse that is frequently smoked (in the crack form), and cocaine-dependent individuals also show aberrant disgust responses [30]. Aberrant disgust responses are also implicated in opiate [31] and internet addiction [32]. The cannabis cessation subreddit also demonstrated outlier status on “disgust” although here it was only observed for the discovery data, suggesting that this aspect of the subjective experience of cannabis cessation may be intensifying over time. The use of “anxiety” was also high in this subreddit, consistent with well-established reports of a link between cannabis use and anxiety [33,34], particularly social anxiety [35-37]. The alcohol cessation subreddit demonstrated outlier status on the “pride” and “gratitude” emotions, the latter was consistent with the well-established role of gratitude in Alcoholics Anonymous (as well as Narcotics Anonymous) programs [19,38,39], and theories linking gratitude with improved outcomes in alcohol use disorder [40]. The smoking subreddit was also highly ranked on the “pride” emotion, although not always exceeding the 95th percentile in representation. In general, relatively less research [41] has been devoted to the role of pride in substance use cessation [19].

Using cosine similarity to quantify shared emotion patterns, specifically of relative occurrence frequencies across emotion words, we showed that the 3 substance cessation subreddits were more emotionally similar to each other than to other subreddits in the sample. Nevertheless, this substance cessation emotion phenotype appeared, to a lesser degree, in other relevant subreddits, for example, those concerned with social anxiety and ADHD. Consistent with this finding, young adults with ADHD are 1.5 times more likely to develop a substance use disorder for alcohol, cigarette smoking, and illicit drugs (for which cannabis is the most common drug of abuse) [42]. One question raised by this phenomenological overlap in emotion profiles is whether the implicated disorders (eg, substance use disorder but also social anxiety and ADHD) share underlying predisposing conditions or genetic factors that affect multiple emotion system traits [43]. This finding may also have important implications for research into potential shared neurobiological correlates as well as policy implications for targeting public health prevention and intervention efforts.

Our abstinence-linked analyses of r/stopdrinking and r/stopsmoking showed a disproportionately high tendency to post within the first 24 hours of abstinence. This finding suggests that, across substances, these first 24 hours may be a special window associated with the greatest inclination to engage with support communities, possibly offering a key intervention point in recovery from addiction. This result contrasts with the culture of some offline support groups, for instance, Cocaine Anonymous, which discourages members from sharing at meetings within the first 24 hours of abstinence [44]. Beyond the first day, a significant drop in the “anxiety” emotion was observed for longer compared to shorter durations of abstinence from alcohol, consistent with prior reports suggesting that anxiety linked to alcohol cessation tends to resolve after about 6 weeks of abstinence [45]. Although not significant following Bonferroni correction, we speculate that the finding of greater “gratitude” emotion with longer alcohol abstinence duration may suggest a cumulative effect of exposure to alcohol cessation support groups, which are strongly influenced by the culture of gratitude at Alcoholics Anonymous. While these results await replication with objectively verified abstinence duration (eg, by using urine drug screens and prospective longitudinal designs), these changes in emotion expression may also be linked to the time course of craving incubation [46,47].

The 4 outlier emotions identified above, as well as the overall substance-preferential emotion pattern identified in our cosine similarity analysis, were robustly replicated across 2 different time periods, pre- and post-COVID-19, during the same season of the year (late autumn to early winter). The transdiagnostic connections to nonsubstance-themed subreddits were also robust, with r/ADHD and r/socialskills surfacing in both pre- and post-COVID-19 results. This degree of replication was especially powerful given that the COVID-19 pandemic was previously shown to have affected posting behavior on Reddit, particularly within support groups, including spikes in posts relating to health anxiety across multiple forums [48]. The reproducibility of these emotion profile properties is therefore suggestive of a stable subjective experience phenotype related to substance cessation.

Limitations

We first acknowledge the absence of independent or objective measures as a limitation to validating these subjective reports. We note that the characteristics of the data set do not permit verification of a clinical diagnosis. Therefore, the degree to which this substance cessation-preferential emotion profile may explain individual variation in disease risk, progression, prognosis (eg, relapse risk), or phase in abstinence remains unknown. Second, our data set is limited by selection effects that bias our findings toward subjects who are English literate, have access to the internet, and are motivated to self-disclose in written form about these potentially sensitive topics. In general, it is important to acknowledge potential bias introduced by the overall demographics of Reddit users, who are predominantly White, male, and college-educated individuals who are 18-29 years of age and living in the United States [49]. Other selection biases include geographic and temporal constraints in the collected data, which may also influence sentiment expression [50,51]. These biases may have excluded many other groups of individuals struggling with substance cessation, and the overlap with participants in laboratory-based studies may be limited. It is also important to note that the substance cessation subreddits were smaller (less than 1 million members) than the subreddits in the comparison sample (all with at least 1 million members), which represented the “general population” control sample. This difference in membership size is an unavoidable potential to confound in comparing control population interests with a relatively niche interest. However, given that the preferential similarity across the substance cessation subreddits did not replicate when using a time word bank, our results argue against a trivial consequence of a difference in membership size. We also cannot exclude the possibility that some emotion similarities identified arose from overlapping membership across substance cessation subreddits and other subreddits such as r/ADHD, although a recent subreddit similarity analysis based only on shared membership identified relationships distinct from our results [52]. Future efforts could use pretrained language models [53] or transformer models [54,55], which could capture context differently than single words. For example, automated tools for NLP can leverage a wider range of semantic and lexical features in speech and text, such as emotional valence and negation of emotion words [56], pronoun use [57], and linguistic complexity [58],

which may provide a richer selection of features for the sentiment analysis. While not directly pertinent for the purposes of this study, which intended to identify the frequency of commonly expressed emotions on substance cessation forums, these concepts could add a level of specificity to the interpretation of the results. Future analyses may also benefit from sampling more deeply on the order of tens of thousands or hundreds of thousands of posts. Nevertheless, our use of a priori selected word banks to identify specific patterns of emotional expression suggests that important emotional features can be captured in these data even at the relatively small scale of 5000 raw posts extracted per subreddit.

Conclusions

In conclusion, our results identify a reproducible subjective experience phenotype linked to a key phase of recovery in addiction: substance cessation. This phenotype showed extensive cross-substance overlap, was emotion specific, and may be partly generalizable to other related nonsubstance mental health conditions. Our approach and results could open a window into assessment and prevention efforts in at-risk substance-naive subjects. For example, our methods are conducive to targeting large populations with a strong presence on web-based social media (eg, adolescents and young adults) as well as other at-risk groups, allowing for reach to those not commonly enrolling in research or treatment. Substance use cessation prevention messages (including a link to available resources) may thus be included in sidebars for the r/ADHD subreddit, and anxiety support resources might be introduced into any of the substance cessation subreddits. Further, results can be used to better target the wording of public service announcements (including those used by social media and related technology), tailored to key time windows within the substance cessation trajectory. For example, interventions addressing anxious feelings may be best targeted at recently abstinent populations, whereas gratitude-laden messaging might be more appropriate for longer-term abstainers. In general, our results support ongoing efforts to characterize potential unique features of language use in populations with addiction, explored using NLP to identify behaviors that characterize substance use disorders or predict longitudinal drug use outcomes. If validated in a population clinically diagnosed with addiction, these results can inform timely prevention and intervention efforts on a large scale.

Acknowledgments

This work was supported by the National Institutes of Health grant R01DA041528 to RZG.

Data Availability

The data sets analyzed during this study are available in Open Science Framework Storage: DOI 10.17605/OSF.IO/FCG8V.

Authors' Contributions

GY and RZG conceptualized this study. GY performed the data acquisition. GY, SGK, and H-ML analyzed the data. GY and SGK wrote the initial paper. GY, SGK, H-ML, and RZG revised this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary information supporting methods and results in the main text.

[\[DOCX File , 1002 KB-Multimedia Appendix 1\]](#)

References

1. Koob GF, Le Moal M. Drug addiction, dysregulation of reward, and allostasis. *Neuropsychopharmacol* 2001;24(2):97-129 [[FREE Full text](#)] [doi: [10.1016/S0893-133X\(00\)00195-0](https://doi.org/10.1016/S0893-133X(00)00195-0)] [Medline: [11120394](#)]
2. George O, Le Moal M, Koob GF. Allostasis and addiction: role of the dopamine and corticotropin-releasing factor systems. *Physiol Behav* 2012;106(1):58-64 [[FREE Full text](#)] [doi: [10.1016/j.physbeh.2011.11.004](https://doi.org/10.1016/j.physbeh.2011.11.004)] [Medline: [22108506](#)]
3. Pandey SC. The gene transcription factor cyclic AMP-responsive element binding protein: role in positive and negative affective states of alcohol addiction. *Pharmacol Ther* 2004;104(1):47-58 [[FREE Full text](#)] [doi: [10.1016/j.pharmthera.2004.08.002](https://doi.org/10.1016/j.pharmthera.2004.08.002)] [Medline: [15500908](#)]
4. Ray LA, Bujarski S, Roche DJO. Subjective response to alcohol as a research domain criterion. *Alcohol Clin Exp Res* 2016;40(1):6-17 [[FREE Full text](#)] [doi: [10.1111/acer.12927](https://doi.org/10.1111/acer.12927)] [Medline: [26727518](#)]
5. Verdejo-García A, Bechara A, Recknor EC, Pérez-García M. Executive dysfunction in substance dependent individuals during drug use and abstinence: an examination of the behavioral, cognitive and emotional correlates of addiction. *J Int Neuropsychol Soc* 2006;12(3):405-415 [doi: [10.1017/s1355617706060486](https://doi.org/10.1017/s1355617706060486)] [Medline: [16903133](#)]
6. Fox HC, Axelrod SR, Paliwal P, Sleeper J, Sinha R. Difficulties in emotion regulation and impulse control during cocaine abstinence. *Drug Alcohol Depend* 2007;89(2-3):298-301 [[FREE Full text](#)] [doi: [10.1016/j.drugalcdep.2006.12.026](https://doi.org/10.1016/j.drugalcdep.2006.12.026)] [Medline: [17276626](#)]
7. Zeiger JS, Haberstick BC, Corley RP, Ehringer MA, Crowley TJ, Hewitt JK, et al. Subjective effects for alcohol, tobacco, and marijuana association with cross-drug outcomes. *Drug Alcohol Depend* 2012;123(Suppl 1):S52-S58 [[FREE Full text](#)] [doi: [10.1016/j.drugalcdep.2012.02.014](https://doi.org/10.1016/j.drugalcdep.2012.02.014)] [Medline: [22445481](#)]
8. Sofuoglu M, DeVito EE, Waters AJ, Carroll KM. Cognitive function as a transdiagnostic treatment target in stimulant use disorders. *J Dual Diagn* 2016;12(1):90-106 [[FREE Full text](#)] [doi: [10.1080/15504263.2016.1146383](https://doi.org/10.1080/15504263.2016.1146383)] [Medline: [26828702](#)]
9. Koob GF. The dark side of emotion: the addiction perspective. *Eur J Pharmacol* 2015;753:73-87 [[FREE Full text](#)] [doi: [10.1016/j.ejphar.2014.11.044](https://doi.org/10.1016/j.ejphar.2014.11.044)] [Medline: [25583178](#)]
10. Volkow ND. The reality of comorbidity: depression and drug abuse. *Biol Psychiatry* 2004;56(10):714-717 [doi: [10.1016/j.biopsych.2004.07.007](https://doi.org/10.1016/j.biopsych.2004.07.007)] [Medline: [15556111](#)]
11. Jones MN, editor. *Big Data in Cognitive Science*. 1st Edition. London, England: Psychology Press; 2016.
12. Liu Y, Whitfield C, Zhang T, Hauser A, Reynolds T, Anwar M. Monitoring COVID-19 pandemic through the lens of social media using natural language processing and machine learning. *Health Inf Sci Syst* 2021;9(1):25 [[FREE Full text](#)] [doi: [10.1007/s13755-021-00158-4](https://doi.org/10.1007/s13755-021-00158-4)] [Medline: [34188896](#)]
13. Wang J, Fan Y, Palacios J, Chai Y, Guetta-Jeanrenaud N, Obradovich N, et al. Global evidence of expressed sentiment alterations during the COVID-19 pandemic. *Nat Hum Behav* 2022;6(3):349-358 [[FREE Full text](#)] [doi: [10.1038/s41562-022-01312-y](https://doi.org/10.1038/s41562-022-01312-y)] [Medline: [35301467](#)]
14. Kim SJ, Marsch LA, Hancock JT, Das AK. Scaling up research on drug abuse and addiction through social media big data. *J Med Internet Res* 2017;19(10):e353 [[FREE Full text](#)] [doi: [10.2196/jmir.6426](https://doi.org/10.2196/jmir.6426)] [Medline: [29089287](#)]
15. Hu M, Benson R, Chen AT, Zhu SH, Conway M. Determining the prevalence of cannabis, tobacco, and vaping device mentions in online communities using natural language processing. *Drug Alcohol Depend* 2021 Nov 01;228:109016 [[FREE Full text](#)] [doi: [10.1016/j.drugalcdep.2021.109016](https://doi.org/10.1016/j.drugalcdep.2021.109016)] [Medline: [34560332](#)]
16. Oram D, Tzilos Wernette G, Nichols LP, Vydiswaran VGV, Zhao X, Chang T. Substance use among young mothers: an analysis of Facebook posts. *JMIR Pediatr Parent* 2018 Dec 04;1(2):e10261 [[FREE Full text](#)] [doi: [10.2196/10261](https://doi.org/10.2196/10261)] [Medline: [31518312](#)]
17. Myslín M, Zhu SH, Chapman W, Conway M. Using Twitter to examine smoking behavior and perceptions of emerging tobacco products. *J Med Internet Res* 2013;15(8):e174 [[FREE Full text](#)] [doi: [10.2196/jmir.2534](https://doi.org/10.2196/jmir.2534)] [Medline: [23989137](#)]
18. Plutchik R. Chapter 1—A general psychoevolutionary theory of emotion. In: Plutchik R, Kellerman H, editors. *Theories of Emotion*. Cambridge, MA: Academic Press; 1980:3-33
19. Anonymous. *Twelve Steps and Twelve Traditions*. 1st Edition. New York: AA World Services; 2002.
20. Claudio C. 5000 most common words (part 1)—vocabulary list. Vocabulary.com. URL: <https://www.vocabulary.com/lists/154147> [accessed 2023-06-26]
21. List of emotions: 53 ways to express what you're feeling. Healthline Media. URL: <https://www.healthline.com/health/list-of-emotions> [accessed 2023-06-26]
22. List of emotions: 271 emotion words (+ PDF). The Berkeley Well-Being Institute. URL: <https://www.berkeleywellbeing.com/list-of-emotions.html> [accessed 2023-06-26]
23. Mikolov T, Chen K, Corrado G, Dean J. Efficient Estimation of Word Representations in Vector Space. ArXiv. Preprint posted online on September 7, 2013 [[FREE Full text](#)] [doi: [10.48550/arXiv.1301.3781](https://doi.org/10.48550/arXiv.1301.3781)]

24. Řehůřek R, Sojka P. Software framework for topic modelling with large corpora. In: Proceedings of LREC 2010 workshop New Challenges for NLP Frameworks. Malta: University of Malta; 2010:46-50
25. Conway FN, Samora J, Brinkley K, Jeong H, Clinton N, Claborn KR. Impact of COVID-19 among people who use drugs: a qualitative study with harm reduction workers and people who use drugs. *Harm Reduct J* 2022;19(1):72 [FREE Full text] [doi: [10.1186/s12954-022-00653-1](https://doi.org/10.1186/s12954-022-00653-1)] [Medline: [35780109](https://pubmed.ncbi.nlm.nih.gov/35780109/)]
26. Cochran JR, Kydd RR, Lee JMJ, Walker N, Consedine NS. Disgust but not health anxiety graphic warning labels reduce motivated attention in smokers: a study of P300 and late positive potential responses. *Nicotine Tob Res* 2018;20(7):819-826 [doi: [10.1093/ntr/ntx158](https://doi.org/10.1093/ntr/ntx158)] [Medline: [29065198](https://pubmed.ncbi.nlm.nih.gov/29065198/)]
27. Clayton RB, Leshner G, Tomko RL, Trull TJ, Piasecki TM. Countering craving with disgust images: examining nicotine withdrawn smokers' motivated message processing of anti-tobacco public service announcements. *J Health Commun* 2017;22(3):254-261 [FREE Full text] [doi: [10.1080/10810730.2016.1268222](https://doi.org/10.1080/10810730.2016.1268222)] [Medline: [28248620](https://pubmed.ncbi.nlm.nih.gov/28248620/)]
28. Clayton RB, Leshner G, Bolls PD, Thorson E. Discard the smoking cues-keep the disgust: an investigation of tobacco smokers' motivated processing of anti-tobacco commercials. *Health Commun* 2017;32(11):1319-1330 [doi: [10.1080/10410236.2016.1220042](https://doi.org/10.1080/10410236.2016.1220042)] [Medline: [27690639](https://pubmed.ncbi.nlm.nih.gov/27690639/)]
29. Li X, Li W, Chen H, Cao N, Zhao B. Cigarette-specific disgust aroused by smoking warning images strengthens smokers' inhibitory control under smoking-related background in Go/NoGo task. *Psychopharmacol (Berl)* 2021;238(10):2827-2838 [doi: [10.1007/s00213-021-05898-5](https://doi.org/10.1007/s00213-021-05898-5)] [Medline: [34169344](https://pubmed.ncbi.nlm.nih.gov/34169344/)]
30. Ersche KD, Hagan CC, Smith DG, Abbott S, Jones PS, Apergis-Schoute AM, et al. Aberrant disgust responses and immune reactivity in cocaine-dependent men. *Biol Psychiatry* 2014 Jan 15;75(2):140-147 [FREE Full text] [doi: [10.1016/j.biopsych.2013.08.004](https://doi.org/10.1016/j.biopsych.2013.08.004)] [Medline: [24090796](https://pubmed.ncbi.nlm.nih.gov/24090796/)]
31. Martin L, Clair J, Davis P, O'Ryan D, Hoshi R, Curran HV. Enhanced recognition of facial expressions of disgust in opiate users receiving maintenance treatment. *Addiction* 2006;101(11):1598-1605 [doi: [10.1111/j.1360-0443.2006.01574.x](https://doi.org/10.1111/j.1360-0443.2006.01574.x)] [Medline: [17034439](https://pubmed.ncbi.nlm.nih.gov/17034439/)]
32. Chen Z, Poon KT, Cheng C. Deficits in recognizing disgust facial expressions and internet addiction: perceived stress as a mediator. *Psychiatry Res* 2017;254:211-217 [FREE Full text] [doi: [10.1016/j.psychres.2017.04.057](https://doi.org/10.1016/j.psychres.2017.04.057)] [Medline: [28476013](https://pubmed.ncbi.nlm.nih.gov/28476013/)]
33. Crippa JA, Zuardi AW, Martín-Santos R, Bhattacharyya S, Atakan Z, McGuire P, et al. Cannabis and anxiety: a critical review of the evidence. *Hum Psychopharmacol* 2009;24(7):515-523 [doi: [10.1002/hup.1048](https://doi.org/10.1002/hup.1048)] [Medline: [19693792](https://pubmed.ncbi.nlm.nih.gov/19693792/)]
34. Cheung JTW, Mann RE, Ialomiteanu A, Stoduto G, Chan V, Ala-Leppilampi K, et al. Anxiety and mood disorders and cannabis use. *Am J Drug Alcohol Abuse* 2010;36(2):118-122 [doi: [10.3109/00952991003713784](https://doi.org/10.3109/00952991003713784)] [Medline: [20337509](https://pubmed.ncbi.nlm.nih.gov/20337509/)]
35. Buckner JD, Zvolensky MJ, Schmidt NB. Cannabis-related impairment and social anxiety: the roles of gender and cannabis use motives. *Addict Behav* 2012;37(11):1294-1297 [FREE Full text] [doi: [10.1016/j.addbeh.2012.06.013](https://doi.org/10.1016/j.addbeh.2012.06.013)] [Medline: [22766487](https://pubmed.ncbi.nlm.nih.gov/22766487/)]
36. Buckner JD, Schmidt NB, Lang AR, Small JW, Schlauch RC, Lewinsohn PM. Specificity of social anxiety disorder as a risk factor for alcohol and cannabis dependence. *J Psychiatr Res* 2008;42(3):230-239 [FREE Full text] [doi: [10.1016/j.jpsychires.2007.01.002](https://doi.org/10.1016/j.jpsychires.2007.01.002)] [Medline: [17320907](https://pubmed.ncbi.nlm.nih.gov/17320907/)]
37. Agosti V, Nunes E, Levin F. Rates of psychiatric comorbidity among U.S. residents with lifetime cannabis dependence. *Am J Drug Alcohol Abuse* 2002;28(4):643-652 [doi: [10.1081/ada-120015873](https://doi.org/10.1081/ada-120015873)] [Medline: [12492261](https://pubmed.ncbi.nlm.nih.gov/12492261/)]
38. Chen G. Does gratitude promote recovery from substance misuse? *Addict Res Theory* 2017;25(2):121-128 [doi: [10.1080/16066359.2016.1212337](https://doi.org/10.1080/16066359.2016.1212337)]
39. Anonymous. *Alcoholics Anonymous: The Story of How Many Thousands of Men and Women Have Recovered From Alcoholism/B-1*. 3rd Revised Edition. New York: Alcoholics Anonymous World Services, Inc; 1976.
40. LaBelle OP, Edelstein RS. Gratitude, insecure attachment, and positive outcomes among 12-step recovery program participants. *Addict Res Theory* 2017;26(2):123-132 [doi: [10.1080/16066359.2017.1333111](https://doi.org/10.1080/16066359.2017.1333111)]
41. Pados E, Kovács A, Kiss D, Kassai S, Kapitány-Fövény M, Dávid F, et al. Voices of temporary sobriety—a diary study of an alcohol-free month in Hungary. *Subst Use Misuse* 2020;55(5):839-850 [FREE Full text] [doi: [10.1080/10826084.2019.1705861](https://doi.org/10.1080/10826084.2019.1705861)] [Medline: [31900020](https://pubmed.ncbi.nlm.nih.gov/31900020/)]
42. Wilens TE, Martelon M, Joshi G, Bateman C, Fried R, Petty C, et al. Does ADHD predict substance-use disorders? A 10-year follow-up study of young adults with ADHD. *J Am Acad Child Adolesc Psychiatry* 2011;50(6):543-553 [FREE Full text] [doi: [10.1016/j.jaac.2011.01.021](https://doi.org/10.1016/j.jaac.2011.01.021)] [Medline: [21621138](https://pubmed.ncbi.nlm.nih.gov/21621138/)]
43. Linnér RK, Mallard TT, Barr PB, Sanchez-Roige S, Madole JW, Driver MN, COGA Collaborators, et al. Multivariate analysis of 1.5 million people identifies genetic associations with traits related to self-regulation and addiction. *Nat Neurosci* 2021;24(10):1367-1376 [FREE Full text] [doi: [10.1038/s41593-021-00908-3](https://doi.org/10.1038/s41593-021-00908-3)] [Medline: [34446935](https://pubmed.ncbi.nlm.nih.gov/34446935/)]
44. Readings from our literature. *Cocaine Anonymous New York*. URL: <https://canewyork.org/literature> [accessed 2021-10-31]
45. Schuckit MA, Hesselbrock V. Alcohol dependence and anxiety disorders: what is the relationship? *Am J Psychiatry* 1994;151(12):1723-1734 [doi: [10.1176/ajp.151.12.1723](https://doi.org/10.1176/ajp.151.12.1723)] [Medline: [7977877](https://pubmed.ncbi.nlm.nih.gov/7977877/)]
46. Li X, Caprioli D, Marchant NJ. Recent updates on incubation of drug craving: a mini-review. *Addict Biol* 2015;20(5):872-876 [FREE Full text] [doi: [10.1111/adb.12205](https://doi.org/10.1111/adb.12205)] [Medline: [25440081](https://pubmed.ncbi.nlm.nih.gov/25440081/)]

47. Parvaz MA, Moeller SJ, Goldstein RZ. Incubation of Cue-Induced Craving in Adults Addicted to Cocaine Measured by Electroencephalography. *JAMA Psychiatry* 2016 Nov 01;73(11):1127-1134 [FREE Full text] [doi: [10.1001/jamapsychiatry.2016.2181](https://doi.org/10.1001/jamapsychiatry.2016.2181)] [Medline: [27603142](https://pubmed.ncbi.nlm.nih.gov/27603142/)]
48. Low DM, Rumker L, Talkar T, Torous J, Cecchi G, Ghosh SS. Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on reddit during COVID-19: observational study. *J Med Internet Res* 2020;22(10):e22635 [FREE Full text] [doi: [10.2196/22635](https://doi.org/10.2196/22635)] [Medline: [32936777](https://pubmed.ncbi.nlm.nih.gov/32936777/)]
49. Barthel M, Stocking G, Holcomb J, Mitchell A. Reddit news users more likely to be male, young and digital in their news preferences. Pew Research Center. 2016. URL: <https://www.pewresearch.org/journalism/2016/02/25/reddit-news-users-more-likely-to-be-male-young-and-digital-in-their-news-preferences/> [accessed 2023-06-23]
50. Mitchell L, Frank MR, Harris KD, Dodds PS, Danforth CM. The geography of happiness: connecting Twitter sentiment and expression, demographics, and objective characteristics of place. *PLoS One* 2013;8(5):e64417 [FREE Full text] [doi: [10.1371/journal.pone.0064417](https://doi.org/10.1371/journal.pone.0064417)] [Medline: [23734200](https://pubmed.ncbi.nlm.nih.gov/23734200/)]
51. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. *PLoS One* 2018;13(6):e0198857 [FREE Full text] [doi: [10.1371/journal.pone.0198857](https://doi.org/10.1371/journal.pone.0198857)] [Medline: [29902270](https://pubmed.ncbi.nlm.nih.gov/29902270/)]
52. Waller I, Anderson A. Quantifying social organization and political polarization in online platforms. *Nature* 2021;600(7888):264-268 [doi: [10.1038/s41586-021-04167-x](https://doi.org/10.1038/s41586-021-04167-x)] [Medline: [34853472](https://pubmed.ncbi.nlm.nih.gov/34853472/)]
53. Shon S, Brusco P, Pan J, Han KJ, Watanabe S. Leveraging pre-trained language model for speech sentiment analysis. ArXiv. Preprint posted online on June 11, 2021 [FREE Full text] [doi: [10.21437/interspeech.2021-1723](https://doi.org/10.21437/interspeech.2021-1723)]
54. Biswas E, Karabulut ME, Pollock L, Vijay-Shanker K. Achieving reliable sentiment analysis in the software engineering domain using BERT. 2020 Presented at: 2020 IEEE International Conference on Software Maintenance and Evolution; September 28-October 02, 2020; Adelaide, SA, Australia p. 162-173 URL: <https://ieeexplore.ieee.org/document/9240599> [doi: [10.1109/icsme46990.2020.00025](https://doi.org/10.1109/icsme46990.2020.00025)]
55. Kumar S, Sumers TR, Yamakoshi T, Goldstein A, Hasson U, Norman KA, et al. Reconstructing the cascade of language processing in the brain using the internal computations of a transformer-based language model. bioRxiv. Preprint posted online on February 23, 2023 [FREE Full text] [doi: [10.1101/2022.06.08.495348](https://doi.org/10.1101/2022.06.08.495348)]
56. Tanana MJ, Soma CS, Kuo PB, Bertagnolli NM, Dembe A, Pace BT, et al. How do you feel? Using natural language processing to automatically rate emotion in psychotherapy. *Behav Res Methods* 2021 Oct;53(5):2069-2082 [FREE Full text] [doi: [10.3758/s13428-020-01531-z](https://doi.org/10.3758/s13428-020-01531-z)] [Medline: [33754322](https://pubmed.ncbi.nlm.nih.gov/33754322/)]
57. Brockmeyer T, Zimmermann J, Kulesa D, Hautzinger M, Bents H, Friederich HC, et al. Me, myself, and I: self-referent word use as an indicator of self-focused attention in relation to depression and anxiety. *Front Psychol* 2015;6:1564 [FREE Full text] [doi: [10.3389/fpsyg.2015.01564](https://doi.org/10.3389/fpsyg.2015.01564)] [Medline: [26500601](https://pubmed.ncbi.nlm.nih.gov/26500601/)]
58. Bilgrami ZR, Sarac C, Srivastava A, Herrera SN, Azis M, Haas SS, et al. Construct validity for computational linguistic metrics in individuals at clinical risk for psychosis: associations with clinical ratings. *Schizophr Res* 2022 Jul;245:90-96 [FREE Full text] [doi: [10.1016/j.schres.2022.01.019](https://doi.org/10.1016/j.schres.2022.01.019)] [Medline: [35094918](https://pubmed.ncbi.nlm.nih.gov/35094918/)]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder

NLP: natural language processing

Edited by A Mavragani; submitted 22.12.22; peer-reviewed by ID Yucel, R Gore; comments to author 12.04.23; revised version received 02.05.23; accepted 09.06.23; published 19.07.23

Please cite as:

Yang G, King SG, Lin HM, Goldstein RZ

Emotional Expression on Social Media Support Forums for Substance Cessation: Observational Study of Text-Based Reddit Posts
J Med Internet Res 2023;25:e45267

URL: <https://www.jmir.org/2023/1/e45267>

doi: [10.2196/45267](https://doi.org/10.2196/45267)

PMID:

©Genevieve Yang, Sarah G King, Hung-Mo Lin, Rita Z Goldstein. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 19.07.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The

complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.