The American Journal of Drug and Alcohol Abuse

Encompassing All Addictive Disorders

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/iada20

Predicting changes in substance use following psychedelic experiences: natural language processing of psychedelic session narratives

David J. Cox, Albert Garcia-Romeu & Matthew W. Johnson

THE AMERICAN JOURNAL OF DRUG AND ALCOHOL ABUSE
https://doi.org/10.1080/00952990.2021.1910830

ORIGINAL ARTICLE

Predicting changes in substance use following psychedelic experiences: natural language processing of psychedelic session narratives

David J. Cox[a,b], Albert Garcia-Romeu[a], and Matthew W. Johnson[a]

[a]Behavioral Pharmacology Research Unit, Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; [b]Department of Data & Analytics, Behavioral Health Center of Excellence, Los Angeles, California, USA

ABSTRACT

Background: Experiences with psychedelic drugs, such as psilocybin or lysergic acid diethylamide (LSD), are sometimes followed by changes in patterns of tobacco, opioid, and alcohol consumption. But, the specific characteristics of psychedelic experiences that lead to changes in drug consumption are unknown.
Objective: Determine whether quantitative descriptions of psychedelic experiences derived using Natural Language Processing (NLP) would allow us to predict who would quit or reduce using drugs following a psychedelic experience.
Methods: We recruited 1141 individuals (247 female, 894 male) from online social media platforms who reported quitting or reducing using alcohol, cannabis, opioids, or stimulants following a psychedelic experience to provide a verbal narrative of the psychedelic experience they attributed as leading to their reduction in drug use. We used NLP to derive topic models that quantitatively described each participant's psychedelic experience narrative. We then used the vector descriptions of each participant's psychedelic experience narrative as input into three different supervised machine learning algorithms to predict long-term drug reduction outcomes.
Results: We found that the topic models derived through NLP led to quantitative descriptions of participant narratives that differed across participants when grouped by the drug class quit as well as the long-term quit/reduction outcomes. Additionally, all three machine learning algorithms led to similar prediction accuracy (~65%, CI = ±0.21%) for long-term quit/reduction outcomes.
Conclusions: Using machine learning to analyze written reports of psychedelic experiences may allow for accurate prediction of quit outcomes and what drug is quit or reduced within psychedelic therapy.

ARTICLE HISTORY
Received 3 October 2020
Revised 26 March 2021
Accepted 26 March 2021

KEYWORDS
Psychedelic treatment; hallucinogens; natural language processing; verbal behavior

Research indicates that psychedelics may effectively aid addiction treatment. Open-label studies administering the serotonin 2A receptor (5-HT2_{AR}) agonist psychedelic psilocybin have shown improvements in tobacco (1,2) and alcohol use disorders (e.g., (3)). A meta-analysis of six studies indicated lysergic acid diethylamide (LSD) showed improved outcomes for alcoholism compared to control conditions (4). Observational studies have reported reductions in substance misuse among religious users of psychedelic-containing plants (e.g., (5,6)).

The psychological mechanisms of psychedelic therapies for addiction remain poorly understood. One theme is that therapeutic outcomes are significantly associated (1,3) or mediated (7, 8) by acute mystical-type effects during the psychedelic experiences. Mystical-type effects involve endorsing qualities including sense of unity, positive mood, transcendence of time and space, and ineffability (9). This is consistent with earlier hypotheses that subjective psychedelic experiences play a critical role in lasting therapeutic benefits (e.g., (10–12)).

However, objectively measuring subjective experiences during a psychedelic session remains difficult. One study analyzed speech during a drug experience to discriminate between ±3,4-methylenedioxymethamphetamine (MDMA, 'ecstasy') and methamphetamine (13) suggesting automated speech analyses using machine learning (ML) could capture subtle alterations in speech assumed indicative of subjective psychedelic experiences. This approach might be viable with MDMA which often involves more talking during acute drug effects compared with 5-HT2_{AR} agonist psychedelics where participants are typically encouraged to minimize real-time description of experiences (14). One alternative method is to have the individual record a narrative of their psychedelic experience after it occurs. Narratives could be analyzed using qualitative methods (e.g.,

CONTACT Matthew W. Johnson mwj@jhu.edu Behavioral Pharmacology Research Unit, Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, 5510 Nathan Shock Drive, Baltimore, MD 21224.

(15,16)). However, qualitative analyses are time-intensive, lack analytic-generalizability, and can be influenced by expectancy (15,17).

One efficient, generalizable, and more objective method to analyze speech is natural language processing (NLP). NLP refers to computational techniques that quantitatively describe human language (18). NLP is increasingly used across many sectors of society (18), including predicting human behavior in psychiatry. For example, NLP has categorized bipolar patients from controls (19); predicted suicidal behavior among hospitalized adolescents (20); and predicted clinical response to psilocybin in patients with depression (21). Coyle and colleagues (22) found NLP of narratives allowed them to discriminate between psychedelic drugs that participants consumed. Further, Zamberlan and colleagues (23) found that speech similarity extracted through NLP correlated with binding affinity profiles of the psychoactive substituted phenethylamines and tryptamines. Researchers have not used NLP to analyze narratives to predict drug use change following psychedelic experiences.

Using survey reports on substance reduction/cessation (24–26), this secondary analysis aimed to determine if NLP of psychedelic narratives could differentiate what drug was reduced/quit, extent of drug use reduction, and how well NLP outputs could predict drug use changes following a psychedelic experience.

Methods

Participants ($N = 1141$) completed an anonymous, 40-min, online survey through SurveyMonkey (https://www.surveymonkey.com) between September 2013 and May 2014. Participants were recruited through social media advertisements distributed at FaceBook.com, Reddit.com, Erowid.org, Shroomery.org, and Maps.org. Recruitment materials solicited participants who had quit or reduced using drugs after a psychedelic experience.

Participants were included if they reported quitting or reducing drug use for any duration of time following use of a 5–$HT2_{AR}$ agonist psychedelic (i.e., psilocybin mushrooms, LSD, morning glory seeds, mescaline, peyote cactus, San Pedro cactus, DMT, and ayahuasca). Participants were grouped by primary drug class they reduced or quit using (alcohol $n = 512$; cannabis n = 272; opioids $n = 195$; stimulants $n = 162$). Participants had to be at least 18 years of age and speak, read, and write English fluently. Participants were not compensated for participation.

The full survey, data collection, and other results are reported in full elsewhere (24–26). Briefly, the survey assessed drug use before and after the psychedelic experience they attributed to cessation/reduction in drug use (hereafter the “reference psychedelic experience”). Participants indicated whether the reference psychedelic experience led to complete abstinence, persistent reduction, or temporary reduction in drug use. Participants endorsing persistent reduction classified current drug use as: ≥1 instance of use/day, ≤1 instance of use/day, ≤1 instance of use/week, or ≤1 instance of use/month. Participants also provided a narrative description of the reference psychedelic experience (no word limits).

Study procedures were approved by the Johns Hopkins University School of Medicine IRB (OHRP Registration 00001656). The study was conducted in accord with the principles within the Declaration of Helsinki and written informed consent was obtained from all participants.

Natural language processing

Topic models

A full description of the computational methods are available online as supplemental information. Psychedelic session narratives were preprocessed using NLTK for Python (27). Preprocessing included removing stop words, lemmatization, and tf-idf transformation (28,29). Following preprocessing of the narratives, we then derived three experimental topic models using latent semantic analysis (LSA; (30)) with singular value decomposition (SVD; (31)).

One topic model used all words and vocabulary from all participants (LSA-All). The second topic model used all words and vocabulary, but only with participants who changed consumption of alcohol (LSA-Alcohol). This allowed us to determine if focusing on a specific drug class led to better differentiation or prediction of drug use outcomes following a psychedelic experience. Alcohol was chosen because it was reduced/quit by the most participants. The LSA-All and LSA-Alcohol topic models may have placed greater weight on the psychedelic drug used in the psychedelic experience rather than language about the experience itself. Therefore, we created a third topic model after removing alcohol-related words (LSA-Scrubbed) from previous research demonstrating significant associations of specific words with craving and alcohol expectancies (32,33). Finally, it is possible simply counting references to specific drugs (e.g., alcohol, cannabis) would be associated with quit outcomes (‘Alcohol Word Count’ model). We, therefore, created a control model where we simply counted the frequency of each alcohol-related word and the sum of all alcohol-related words used by each participant.

The four topic models created quantitative descriptions of each participant’s narrative. For each model, each narrative was described using vectors with n values, where each value was the number associated with a specific topic, and n was the number of topics for that model. For the Alcohol Word Count model, each narrative was described using 32 numbers– 31 were the individual alcohol-related word counts and 1 number was the sum of all alcohol-related words used in their narrative.

Predicting quit outcomes

We next used the output from the topic models as inputs to three different supervised ML algorithms (Table 1). The ML algorithms were trained using topic model results from a randomly selected 75% of participants. Once trained, we used those models to predict outcomes for the remaining 25%. Because the specific random selection of participants can alter model performance, we calculated model metrics over 1000 different random proportions of participants used for training. This provides a likely distribution of successful predictions if the model were used with novel narratives. All ML algorithms were implemented using scikit-learn version 0.21.1 (29) in Python. The three algorithms tested were *k*-nearest neighbors, naïve Bayes classifier, and a random forest classifier (34).

Data analysis

We used Python to conduct 10 planned sets of ANOVAs spanning three questions. Bonferroni adjustments were used within each set of ANOVAs to maintain a family-wise error rate less than 0.05. Follow–up pairwise comparisons were conducted where appropriate. One question was whether the topic models differed as a function of what drug participants reduced/quit using (alcohol, cannabis, opioids, or stimulants). The second question was whether the topic models differed based on quit outcomes. The third question was whether the topic models would predict quit outcomes. To answer this question, three ANOVAs were conducted. Each ANOVA compared model metrics of all four topic models for one supervised ML algorithm.

Table 1. Summary of equations used to calculate coherence scores, *k*-nearest neighbors, naïve Bayes classifier, and the random forest classifier.

Algorithm Name	Equation
Coherence	$Coherence(V) = \sum_{(v_i,v_j)\in V} score(v_i, v_j, \varepsilon)$
k-Nearest Neighbors	$J = \sum_{i=1}^{k}\sum_{j=1}^{n}\left(\lvert\lvert x_i - v_j\rvert\rvert\right)^2 = 1,$
Naïve Bayes	$\hat{\theta}_{yi} = \frac{\sum_{x\in T} x_i + \alpha}{\sum_{i=1}^{n} N_{yi} + \alpha n}$
Random Forest	$ni_j = w_j C_j - w_{left(j)} C_{left(j)} - w_{right(j)} C_{right(j)}$

Coherence: V is a set of words describing the topic and ε is a smoothing factor that guarantees that *score* returns a real number. *k*-nearest neighbors: $\lvert\lvert x_i - v_j\rvert\rvert$ is the Euclidean distance between a point, x_i, and a centroid, v_j, iterated over all k points in the i^{th} group, for all n groups. Naïve Bayes: $\sum_{x\in T} x_i$ refers to the number of times feature i appears in a sample of class y in the training set T; $\sum_{i=1}^{n} N_{yi}$ is the total count of all features for class y; and α are smoothing priors. Random Forest: ni_j is the importance of node j; w_j is the weighted number of samples reaching node j; C_j is the impurity value of node j calculated by the equation $\sum_{i=1}^{C} f_i(1-f_i)$ where f_i is the frequency of group i at a node and C is the number of unique groups; and the *left(j)* and *right(j)* refer to the weighted number of samples (w) and impurity values (C) associated with the left and right children split off node j.

Results

Figure 1 shows coherence scores for LSA topic models with topic numbers ranging 2–100. Higher coherence scores indicate better quantitative description of the narratives. For LSA-All, the two highest coherence scores were 3 and 6 topics. For LSA-Scrubbed, the highest coherence score was six topics. And, for LSA–Alcohol, the top 1-3 coherence scores were 2, 3, and 9 topics, respectively. Thus, 6, 6, and 9 topics were chosen for further analysis for LSA-All, LSA-Scrubbed, and LSA-Alcohol topic models, respectively.

Table 2 shows the top 10 words by weight for each topic from the topic models. The first word listed had the greatest weight with remaining words listed in decreasing weight. A '-' before a word indicates the word is multiplied as negative when deriving the vector for that topic. For each topic, every word in the entire vocabulary is assigned a weight on a continuous scale. Thus, the same word can be ranked highly in multiple topics (e.g., 'alcohol', 'experience') and every word receives a weight for all topics in every model.

Differentiating drug quit or reduced

Table 3 shows ANOVAs asking whether the LSA-All topic model differentiates what drug participants reduced/quit using. The mean (median) number of words per narrative were 142 (91), 120 (83), 160 (101), and 111 (70) for 'alcohol', 'cannabis, 'opioids', and 'stimulants', respectively. We observed statistically significant differences for all six topics in the LSA-All model with Topic 2 resulting in the largest effect size (η^2 = 0.45). Follow-up comparisons using Tukey HSD showed the LSA-All topic model resulted in quantitative descriptions of psychedelic session narratives that differed between all drugs (i.e., all pairwise comparisons had one topic resulting in statistically significant differences; table available in Supplemental Information). The pairwise comparison with the least number of significant differentiating topics (two) was between the opioid and stimulant groups. Three pairwise comparisons resulted in a significant difference across all six topics. These were between participants who reduced/quit alcohol and opioids, cannabis

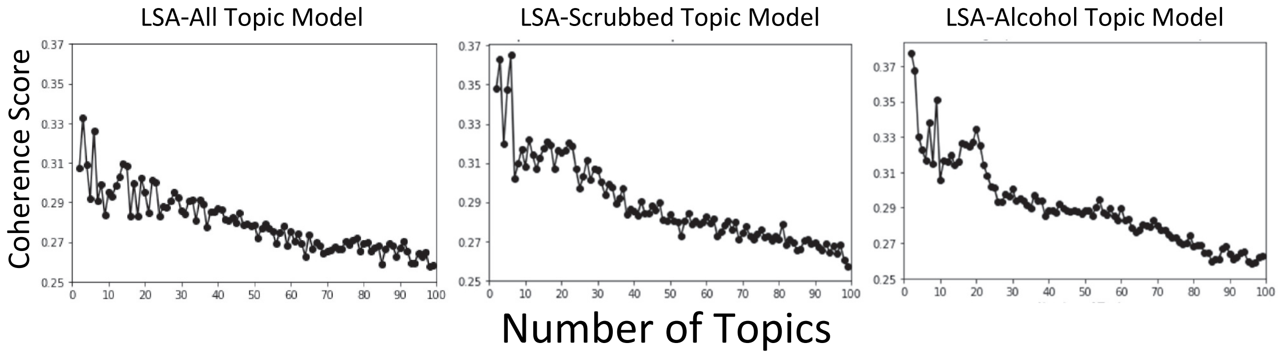

Figure 1. Coherence scores for each topics ranging from 1-to-100 for the Latent Semantic Analysis (LSA) model derived from all participants and using all words (LSA-ALL; left panel), the LSA model derived from all participants but with alcohol-related words scrubbed out (LSA-Scrubbed; middle panel), and the LSA model derived using only participants who quit using alcohol (LSA-Alcohol; right panel).

Table 2. Top 10 words from each topic from each LSA topic model in order of computed weight. *In an abundance of caution regarding confidential participant information we have redacted this proper name, which was used in a single written narrative.

Topic	LSA-All	LSA-Scrubbed	LSA-Alcohol
1	experience, time, take, drink, use, life, go, like, trip, felt	experience, time, take, use, life, go, like, felt, trip, LSD	drink, experience, alcohol, time, take, go, like, life, felt, feel
2	-experience, -drink, -alcohol, go, -use, get, like, trip, smoke, felt	experience, use, -go, -get, LSD, -like, -trip, -say, -think, -would	drink, alcohol, -say, -go, -like, -fall, -would, -felt, experience, -[NAME]*
3	-drink, experience, -alcohol, use, -go, LSD, -get, life, -would, cannabis	smoke, -experience, use, cannabis, trip, LSD, -felt, -like, get, -say	experience, -drink, felt, -go, god, -would, -get, -fall, body, -[NAME]*
4	smoke, -experience, use, cannabis, trip, LSD, -felt, -like, -say, get	-use, -would, smoke, felt, experience, take, -time, -fall, -years, -LSD	experience, -j, would, -take, -start, fall, -mushroom, [NAME]*, -felt, -like
5	-use, -would, take, experience, felt, smoke, -time, -fall, -years, feel	take, -smoke, -cannabis, LSD, -experience, -felt, get, drug, dose, start	alcohol, -experience, -go, -take, felt, years, -friends, -drink, -really, start
6	take, -smoke, LSD, -cannabis, -experience, -felt, drug, get, dose, heroin	felt, use, -experience, -LSD, -trip, like, drug, feel, -go, life	-feel, -drink, LSD, use, alcohol, -god, -felt, -like, take, trip
7	–	–	drug, like, life, really, -j, light, mushrooms, antidepressants, cylinder, shin
8	–	–	time, -drink, god, -like, -felt, -experience, use, get, love, smoke
9	–	–	-god, -alcohol, time, LSD, feel, -one, things, think, -us, -go

Table 3. Results of ANOVAs comparing drug that was quit using topics from the LSA-All topic model. ***Bold and italics*** = significant using Bonferonni corrected alpha threshold of 0.008.

Topic	*dfs*	*F*	*p*	η^2
1	3, 1137	13.01	***<.001***	0.033
2	3, 1137	308.99	***<.001***	0.449
3	3, 1137	68.04	***<.001***	0.152
4	3, 1137	13.67	***<.001***	0.035
5	3, 1137	23.47	***<.001***	0.058
6	3, 1137	41.59	***<.001***	0.099

and opioids, and cannabis and stimulants. In sum, all topics from LSA-All differentiated what drug each participant reduced/quit following a psychedelic experience.

Table 4 shows ANOVAs asking whether the LSA-Scrubbed topic model differentiates what drug each participant reduced/quit using. We observed statistically significant differences for five of six topics with Topic 2 resulting in the largest effect size (η^2 = 0.29). Follow-up comparisons using Tukey HSD showed the LSA-Scrubbed topic model could differentiate 5 of 6 pairwise comparisons. One pairwise comparison involved significant differences across all 6 topics (cannabis and stimulants). The one undifferentiated pairwise comparison was between participants who reduced/quit using alcohol and who reduced/quit using stimulants. In sum, the LSA-Scrubbed topic model differentiated what drug each participant reduced/quit following a psychedelic experience except between alcohol-stimulant groups.

Table 5 shows the ANOVAs asking whether the count of alcohol-related words differentiated what drug participants reduced/quit using. We observed statistically significant differences in the frequency that three

Table 4. Results of ANOVAs comparing drug that was quit using topics from the LSA-Scrubbed topic model. ***Bold and italics*** = significant using Bonferonni corrected alpha threshold of 0.008.

Topic	*dfs*	*F*	*p*	η^2
1	3, 1137	9.83	***<.001***	0.025
2	3, 1137	152.82	***<.001***	0.287
3	3, 1137	7.59	***<.001***	0.020
4	3, 1137	10.71	***<.001***	0.027
5	3, 1137	31.66	***0.001***	0.077
6	3, 1137	3.31	.02	0.009

alcohol-related words were used ('alcohol', 'beer', and 'drink'). We also observed statistically significant differences in the sum of alcohol-related words. The word 'alcohol' resulted in the largest effect size (η^2 = 0.26), followed by the sum of all alcohol-related words (η^2 = 0.22), then 'drink' (η^2 = 0.18), and 'beer' (η^2 = 0.03). Follow-up multiple comparisons showed the alcohol group used more total alcohol words than the cannabis, opioids, and stimulants groups, and more of the three specific alcohol-related words above. All other pairwise comparisons involving non-alcohol participants were not statistically different for any single alcohol-related words nor for the sum of all alcohol-related words. In sum, the Alcohol-Related-Word-Count topic model differentiated participants who reduced/quit alcohol following a psychedelic experience from those who reduced/quit a different drug class.

Table 5. Results of ANOVAs comparing drug that was quit using the alcohol-related word counts. – indicates the word was not used. ***Bold and italics*** = significant using Bonferonni corrected alpha threshold of 0.002.

Word	*dfs*	*F*	*p*	η^2
SUM	3, 1137	108.80	***<.001***	0.223
Alcohol	3, 1137	130.07	***<.001***	0.256
Ale	3, 1137	1.67	.17	0.004
Bar	3, 1137	1.72	.16	0.005
Beer	3, 1137	10.91	***<.001***	0.028
Bitter	3, 1137	0.41	.75	0.001
Booze	3, 1137	2.48	.060	0.006
Bourbon	3, 1137	0.41	.75	0.001
Brandy	3, 1137	–	–	–
Brew	3, 1137	1.71	.16	0.004
Brewery	3, 1137	0.41	.75	0.001
Cider	3, 1137	0.82	.48	0.002
Cocktail	3, 1137	–	–	–
Drink	3, 1137	85.06	***<.001***	0.183
Drunk	3, 1137	2.48	.060	0.006
Gin	3, 1137	0.88	.45	0.002
Lager	3, 1137	–	–	–
Liquor	3, 1137	0.85	.47	0.002
Liqueur	3, 1137	0.41	.75	0.001
Mead	3, 1137	0.41	.75	0.001
Port	3, 1137	2.85	.036	0.007
Pub	3, 1137	1.13	.33	0.003
Rum	3, 1137	2.38	.068	0.006
Saloon	3, 1137	–	–	–
Scotch	3, 1137	0.41	.75	0.001
Sherry	3, 1137	–	–	–
Spirits	3, 1137	–	–	–
Stout	3, 1137	–	–	–
Tavern	3, 1137	–	–	–
Vodka	3, 1137	2.23	.083	0.006
Wine	3, 1137	2.85	.036	0.007
Whiskey	3, 1137	–	–	–
Whiskey	3, 1137	0.82	0.48	0.002

Differentiating quit outcome groups

Table 6 shows the ANOVAs asking whether the LSA-All topic model differentiates reduction/quit outcome groups. The mean (median) number of words per narrative were 137 (84), 138 (88), 98 (63), 146 (84), 99 (80), and 174 (129) for 'stopped completely', 'reduced greatly', 'reduced somewhat', 'stopped but returned to pre-psychedelic rates', 'reduced but returned to pre-psychedelic rates', and 'other', respectively. We observed statistically significant differences for 4 of the 6 topics with Topic 6 showing the largest effect size (η^2 = 0.039). Follow-up multiple comparisons showed the LSA-All topic model differed for nine of the 15 possible pairwise comparisons (table available in Supplemental Information). The six comparisons the LSA-All model did not differentiate were as follows: 'stopped completely' from 'stopped but returned to pre-psychedelic rates'; 'stopped completely' from 'other'; 'reduced greatly' from 'reduced but returned to pre-psychedelic rates'; 'reduced somewhat' from 'reduced but returned to pre-psychedelic rates'; 'stopped but returned to pre-psychedelic rates' from 'reduced but returned to pre-psychedelic rates'; and 'stopped but returned to pre-psychedelic rates' from 'other'. The pairwise comparisons with the most differing topics between the two groups (four) were between 'stopped completely' and 'reduced greatly'.

Table 7 shows the ANOVAs asking whether the LSA-Scrubbed topic model differentiates reduction/quit outcome groups. We observed statistically significant differences in the LSA-Scrubbed model with Topic 5 resulting in the largest effect size (n^2 = 0.03). Follow-up multiple comparisons showed the LSA-Scrubbed topic differentiated 10 of 15 pairwise comparisons. The pairwise comparisons the LSA-Scrubbed did not differentiate were as follows: 'stopped completely' from 'other'; 'reduced somewhat' from 'reduced but returned to pre-psychedelic rates'; 'stopped but returned to pre-psychedelic rates' from 'reduced but returned to pre-psychedelic rates'; 'stopped but returned to pre-psychedelic rates' from 'other'; and 'reduced but returned to pre-psychedelic rates' from 'other'. Two pairwise comparisons had the most topics differing between groups (two). These were between 'stopped completely' and 'reduced somewhat', and between 'reduced somewhat' and 'other'.

Table 8 shows the ANOVAs asking whether the LSA-Alcohol topic model differentiates reduction/quit outcome groups. No topics derived in the LSA-Alcohol model differed between the outcome groups when using Bonferroni corrected alphas across the family of ANOVAs.

Table 6. Results of ANOVAs comparing quit outcomes using topics from the LSA-All topic model. ***Bold and italics*** = significant using Bonferonni corrected alpha threshold of 0.008.

Topic	*dfs*	*F*	*p*	η^2
1	5, 1135	3.54	***.004***	0.015
2	5, 1135	7.70	***<.001***	0.033
3	5, 1135	8.81	***<.001***	0.037
4	5, 1135	1.68	.137	0.007
5	5, 1135	2.74	.018	0.012
6	5, 1135	9.29	***<.001***	0.039

Table 7. Results of ANOVAs comparing quit outcomes using topics from the LSA-Scrubbed topic model. ***Bold and italics*** = significant using Bonferonni corrected alpha threshold of 0.008.

Topic	*dfs*	*F*	*p*	η^2
1	5, 1135	2.60	.023	0.011
2	5, 1135	7.64	***<.001***	0.033
3	5, 1135	2.73	.018	0.012
4	5, 1135	1.84	.103	0.008
5	5, 1135	8.30	***<.001***	0.034
6	5, 1135	4.25	***<.001***	0.019

Table 8. Results of ANOVAs comparing quit outcomes using topics from the LSA-Alcohol topic model. ***Bold and italics*** = significant using Bonferonni corrected alpha threshold of 0.006.

Topic	*dfs*	*F*	*p*	η^2
1	5, 506	2.59	.025	0.025
2	5, 506	1.39	.228	0.014
3	5, 506	0.71	.618	0.007
4	5, 506	0.93	.459	0.009
5	5, 506	2.26	.048	0.022
6	5, 506	0.92	.470	0.009
7	5, 506	1.91	.091	0.019
8	5, 506	0.71	.615	0.007
9	5, 506	0.74	.597	0.007

Table 9. Results of ANOVAs comparing quit outcomes using the count of alcohol-related words. – indicates the word was not used. Bold and italics = significant using Bonferonni corrected alpha threshold of 0.002.

Word	*dfs*	*F*	*p*	η^2
SUM	5, 1135	6.19	***<.001***	0.027
Alcohol	5, 1135	7.46	***<.001***	0.032
Ale	5, 1135	2.70	.020	0.012
Bar	5, 1135	0.69	.63	0.003
Beer	5, 1135	2.38	.037	0.010
Bitter	5, 1135	0.28	.92	0.001
Booze	5, 1135	0.56	0.73	0.002
Bourbon	5, 1135	0.28	.92	0.001
Brandy	5, 1135	–	–	–
Brew	5, 1135	0.31	.91	0.001
Brewery	5, 1135	0.28	.92	0.001
Cider	5, 1135	0.56	.73	0.002
Cocktail	5, 1135	–	–	–
Drink	5, 1135	6.03	***<.001***	0.026
Drunk	5, 1135	0.52	0.76	0.002
Gin	5, 1135	0.96	.44	0.004
Lager	5, 1135	–	–	–
Liquor	5, 1135	0.49	.79	0.002
Liqueur	5, 1135	1.96	.082	0.009
Mead	5, 1135	0.45	.81	0.002
Port	5, 1135	1.68	.14	0.007
Pub	5, 1135	0.13	.99	0.001
Rum	5, 1135	1.16	.32	0.005
Saloon	5, 1135	–	–	–
Scotch	5, 1135	0.28	.92	0.001
Sherry	5, 1135	–	–	–
Spirits	5, 1135	–	–	–
Stout	5, 1135	–	–	–
Tavern	5, 1135	–	–	–
Vodka	5, 1135	1.52	.18	0007
Wine	5, 1135	0.83	.53	0.004
Whiskey	5, 1135	–	–	–
Whiskey	5, 1135	0.56	.73	0.002

Table 9 shows the ANOVAs asking whether the count of alcohol-related words differentiates reduction/quit outcome groups. We observed statistically significant differences in the frequency that 'alcohol' and 'drink' were used. We also observed a statistically significant difference in the sum of alcohol-related words used. Use of 'alcohol' again resulted in the largest effect size (η^2 = 0.032) followed by the sum of alcohol-related words (η^2 = 0.027), and the word 'drink' (η^2 = 0.026). Follow-up multiple comparisons showed the 'reduced greatly' group used the words 'alcohol' and 'drink' more than the 'stopped completely', 'reduced somewhat', and the 'stopped but returned to pre-psychedelic rates' groups. The 'reduced greatly' group also used more total alcohol-related words than the 'stopped completely' and 'reduced somewhat' groups. All other pairwise comparisons were not statistically different.

Predicting outcomes with ML

For all prediction accuracy analyses, we note that random prediction would be accurate in approximately 17% percent of cases (i.e., 100% divided by 6 possible outcomes). Figure 2 shows prediction accuracy for the *k*-nearest neighbor algorithm. We observed a statistically significant difference in prediction accuracies between the four topic models (F(3, 119996) = 29940.76; $p < .001$; $\eta^2 = 0.43$). Follow-up pairwise comparisons indicated prediction accuracies for all topic models were statistically different from all other topic models ($p < .015$ for all comparisons). LSA-Alcohol topic model led to the highest prediction accuracy with median (SD, max) prediction accuracy of 47% (5%, 63%). Although statistically different, LSA-All, LSA-Scrubbed, and Alcohol-Related-Word-Count topic models led to practically similar median (SD, max) accuracies of 39% (4%, 50%), 38% (3%, 48%), and 38% (4%, 48%), respectively.

Figure 3 shows prediction accuracies for the naïve Bayes Bernoulli algorithm. We observed a statistically significant difference in prediction accuracies between topic models (F(3, 3999) = 3701.95; $p < .001$; $\eta^2 = 0.74$). Follow-up pairwise comparisons indicated the prediction accuracies were statistically different between all topic models ($p < .001$ for all comparisons) except between the LSA-Scrubbed and LSA-All models ($t_{1998} = 0.50$; $p = .62$). LSA-Alcohol topic model had the highest

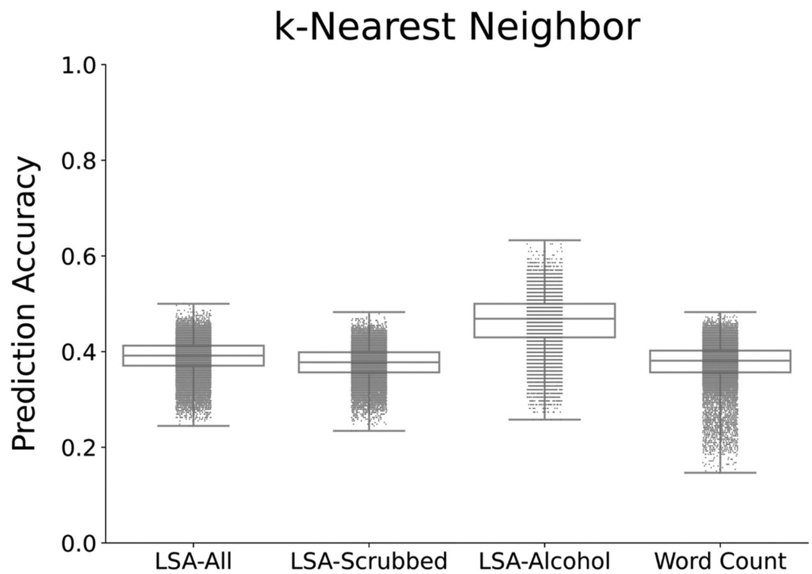

Figure 2. Boxplot overlain with a stripplot of the accuracy scores using the quantitative descriptions of participant psychedelic session narratives as the input for the *k*–nearest neighbor machine learning algorithm. Individual markers in the stripplots are an accuracy score for a single combination of neighbors (1–30) and random states (1-1000). LSA stands for latent semantic analysis.

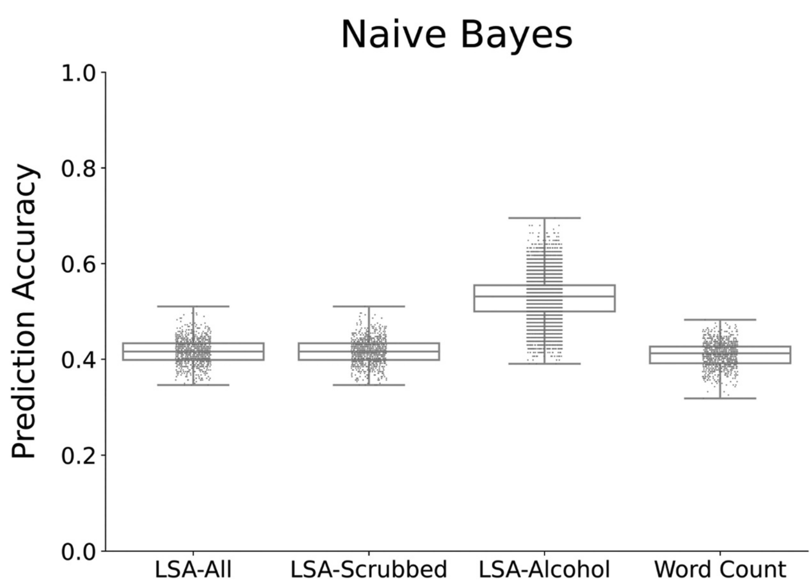

Figure 3. Boxplot of accuracy scores using the quantitative descriptions of participant psychedelic session narratives as the input for the naïve Bayes Bernoulli classifier machine learning clustering algorithm for each random state. LSA stands for latent semantic analysis.

median (SD, max) prediction accuracy of 52% (4%, 68%). The LSA-All, LSA-Scrubbed, and Alcohol-Related-Word -Count topic models again had practically similar median (SD, max) prediction accuracies of 42% (3%, 51%), 42% (3%, 51%), and 41% (3%, 48%), respectively.

Figure 4 shows prediction accuracies for the random forest algorithm. We observed a statistically significant difference in prediction accuracies between the four topic models (F(3, 3996) = 3274.1; $p < .001$; $\eta^2 = 0.72$). Follow-up pairwise comparisons indicated the prediction accuracies were statistically different between all topic models ($p < .001$ for all comparisons) except between the LSA-Scrubbed and Alcohol-Related-Word-Count topic models ($t_{999} = 0.85$; $p = .39$). LSA-Alcohol led to the highest median (SD, max) prediction accuracy of 50% (4%, 63%). The LSA-All, LSA-Scrubbed, and Alcohol-Related-Word-Count topic models resulted in practically similar prediction accuracies with median

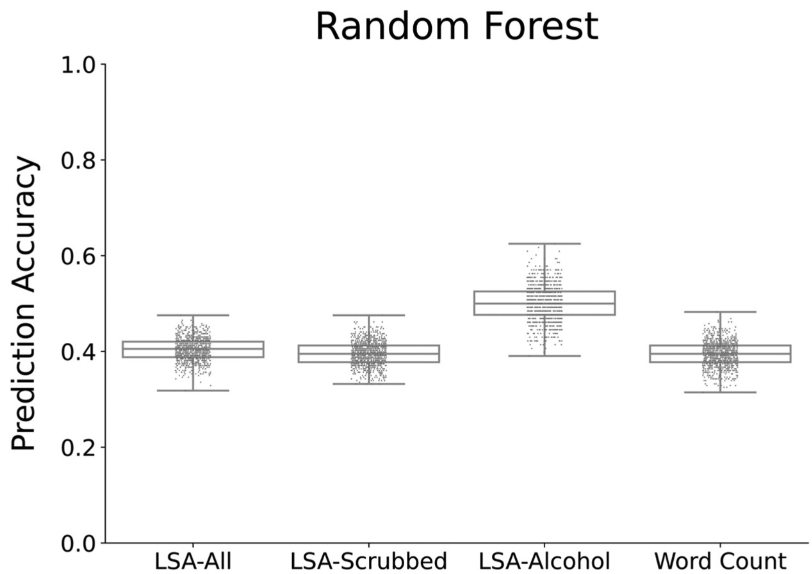

Figure 4. Boxplot of accuracy scores using the quantitative descriptions of participant psychedelic session narratives as the input for the random forest machine learning clustering algorithm for each random state. LSA stands for latent semantic analysis.

(SD, max) prediction accuracies of 41% (2%, 48%), 40% (2%, 48%), and 40% (3%, 48%), respectively.

Discussion

We recruited 1141 individuals who reported decreasing rates of drug consumption following a psychedelic experience. We derived four topic models using NLP, three experimental and one control. Each topic model quantitatively described the verbal descriptions of the psychedelic experience preceding each participant's drug consumption change. We then used each topic model as input to three supervised ML algorithms to predict drug use outcomes. Several notable results emerged. First, the LSA-All and LSA-Scrubbed topic models outperformed the control model in differentiating which drug class participants reduced/quit using. Second, the LSA-All and LSA-Scrubbed topic models outperformed the LSA-Alcohol and Alcohol-Related-Word-Count topic models in differentiating quit outcomes. Third, using the LSA-Alcohol topic model as input to the ML algorithms led to the highest prediction accuracy of quit outcomes. Each finding is discussed.

The LSA-All, LSA-Scrubbed, and Alcohol-Related-Word-Count topic models each differentiated at least two drug classes. The LSA-All model resulted in the largest effect size and differentiated all pairwise comparisons of drug classes. Similar performance between the LSA-All and LSA-Scrubbed topic models suggests each model picked up differences in verbal descriptions apart from alcohol-specific language. This is further supported by results from the control model. The control model resulted in a lower effect size than LSA models and only differentiated participants who reduced/quit alcohol from other drug classes. If the goal of using NLP on written reports of psychedelic experiences is to determine which drugs a person might quit, deriving topic models appears worth the effort compared to simply counting drug-related words.

LSA-All, LSA-Scrubbed, LSA-Alcohol, and Alcohol-Related-Word-Count topic models each differentiated at least two reduction/quit outcomes. Importantly, each model differentiated participants who 'stopped completely' from at least one category of participants who reduced drug use. LSA-All had the largest effect size although all effect sizes were relatively similar and small (all η^2 were 0.04 or 0.03). These results indicate the LSA topic models picked up on differences in psychedelic experience narratives beyond alcohol-related words.

The LSA-All, LSA-Scrubbed, LSA-Alcohol, and Alcohol-Related-Word-Count topic models differentiated 9, 10, 1, and 3 pairwise comparisons in quit outcomes, respectively. Better performance of LSA-All and LSA-Scrubbed models than the LSA-Alcohol model suggests a larger corpus across multiple drug classes resulted in better quantitative descriptions of psychedelic experiences relative to reduction/quit outcomes. Additionally, better performance of LSA-All and LSA-Scrubbed than the control model indicates that deriving topic models seems worth the extra effort if the goal is to understand how differences in a psychedelic experience might be followed by changes in drug use.

Each topic model differentiated the drug class participants reduced/quit and long-term participant quit outcomes (with the possible exception of LSA-Alcohol). We also used supervised ML to determine if LSA-All could predict long-term quit outcomes. Across all algorithms, all topic models resulted in prediction accuracies significantly better than chance. This suggests analyzing psychedelic experience narratives using ML may allow prediction of long-term drug use change following a psychedelic experience.

LSA-Alcohol resulted in the highest prediction accuracies. If the goal of using ML to analyze verbal reports of psychedelic experience narratives is to predict quit outcomes, focusing specifically within a drug class may be the best approach. This contrasts with the findings above that using the larger corpus from all participants led to the greatest difference in quantitative descriptions between reduction/quit outcome groups. Nevertheless, polysubstance use is seemingly the rule, rather than the exception, for individuals with substance use disorder (e.g., (35–37)). Thus, individuals who participate in psychedelic therapy to reduce drug use may consume multiple drugs and the primary drug reduced/quit following a psychedelic experience may be unknown from the outset. In these instances, LSA-All and LSA-Scrubbed led to prediction accuracies significantly better than chance and only slightly worse than the LSA-Alcohol topic model.

All three algorithms led to similar prediction accuracies. Naïve Bayes Bernoulli led to the highest prediction accuracy (68%) compared to the *k*-nearest neighbors (63%) and the random forest (63%) algorithms using LSA-Alcohol as input. However, prediction accuracies were similar for the *k*-nearest neighbors, naïve Bayes Bernoulli, and random forest classifiers using LSA-All (50%, 51%, and 48%, respectively) and LSA-Scrubbed (48%, 51%, and 48%, respectively) as input. Each algorithm involves differing degrees of complexity and computational resources. This study indicates the least complex and resource-intensive algorithm (i.e., *k*-nearest neighbors) led to similar prediction accuracies as the other algorithms. Future research will have to determine the conditions under which more complex and resource-intensive algorithms outperform the *k*-nearest neighbors algorithm such that the tradeoff is worthwhile.

There are several limitations. First, accuracy of the narratives may have been impacted by the delay between the psychedelic experience and recall of the experience. It is possible characteristics of the experience were left out of the narratives, characteristics from a different experience were included in the narrative, or characteristics were incorrectly remembered. Nevertheless, the written reports provided led to quantitative descriptions of the psychedelic experience that differed based on drug class reduced/quit, long–term quit outcomes, and that allowed for prediction of quit outcomes better than chance.

A second limitation is that participants were only those who reduced/quit using a drug following a psychedelic experience and were motivated to talk about it. This population likely differs from those with psychedelic experiences but did not reduce/quit using drugs or are not motivated to talk about it. Therefore, the topic models and prediction models from this study would likely not generalize to other populations. Future research should determine if the same procedures used here would result in similar findings for other populations.

Conclusion

Past research indicates some psychedelic experiences are followed by changing drug consumption including alcohol (3,8,38–41), opioids (12,42), and tobacco (1,2,26,43). Participants in the present study also reported reducing consumption of cannabis and stimulants. Despite a growing body of research indicating reductions in substance use following a psychedelic experience, the mechanisms leading to behavior change are unknown. Past research indicates specific subjective qualities during psychedelic experiences are associated with greater likelihood of behavior change (e.g., (1–3,7,8,44)). The topic models derived here identified differences in reported psychedelic experiences between participants when grouped by the drug class reduced or quit. Past researchers have found that NLP is an efficient, generalizable, and more objective method for categorizing (19,22,23) and predicting behavior (20,21).

The results of the present study extend this literature in three ways. First, by demonstrating NLP can discriminate between which drugs people may reduce or quit using. Second, by demonstrating NLP can discriminate the extent that individuals reduce or quit using specific drugs. Lastly, NLP outputs can be used as inputs to ML to predict quit outcomes following a psychedelic experience. In total, this study indicates that NLP of written psychedelic session narratives and the subsequent use of supervised ML algorithms are analytic tools that can help researchers and clinicians identify the level of support each person may need during psychedelic treatment for substance use disorder.

Acknowledgements

MWJ is in advisory relationships with the following organizations regarding the medical development of psychedelics and related compounds: AWAKN Life Sciences Inc., Beckley Psychedelic Ltd., Entheogen Biomedical Corp., Field Trip Psychedelics Inc., Mind Medicine Inc., Otsuka Pharmaceutical Development & Commercialization Inc., Silo Pharma, Inc.

Funding

Support for MWJ and AGR provided the Johns Hopkins Center for Psychedelic and Consciousness Research funded by Tim Ferriss, Matt Mullenweg, Craig Nerenberg, Blake Mycoskie, and the Steven and Alexandra Cohen Foundation. Support for DJC was provided by National Institute on Drug Abuse Grant [T32DA07209].

ORCID

David J. Cox https://orcid.org/0000-0003-4376-2104
Albert Garcia-Romeu https://orcid.org/0000-0003-2182-1644
Matthew W. Johnson https://orcid.org/0000-0001-7068-0513

References

1. Garcia-Romeu A, Griffiths RR, Johnson MW. Psilocybin occasioned mystical experiences in the treatment of tobacco addiction. Curr Drug Abuse Rev. 2015;7:157–64. doi:10.2174/1874473708666150107121331.
2. Johnson MW, Garcia-Romeu A, Cosimano MP, Griffiths RR. Pilot study of the 5-HT2AR agonist psilocybin in the treatment of tobacco addiction. J Psychopharmacol. 2014;28:983–92. doi:10.1177/0269881114548296.
3. Bogenschutz MP, Forcehimes AA, Pommy JA, Wilcox CE, Barbosa PCR, Strassman RJ. Psilocybin-assisted treatment for alcohol dependence: a proof-of-concept study. J Psychopharmacol. 2015;29:289–99. doi:10.1177/0269881114565144.
4. Krebs TS, Johansen PØ. Over 30 million psychedelic users in the United States. F1000Res. 2013;2:98. doi:10.12688/f1000research.2-98.v1.
5. Albaugh BJ, Anderson PO. Peyote in the treatment of alcoholism among American Indians. Am J Psychiatry. 1974;131:1247–50. doi:10.1176/ajp.131.11.1247.
6. Fabregas JM, Gonzalez D, Fondevila S, Cutchet M, Fernandez X, Barbosa PCR, Alcázar-Córcoles MÁ, Barbanoj MJ, Riba J, Bouso JC, et al. Assessment of addiction severity among ritual users of ayahuasca. Drug Alcohol Depend. 2010;111:257–61.
7. Griffiths RR, Johnson MW, Carducci MA, Umbricht A, Richards WA, Richards BD, Cosimano MP, Klinedinst MA. Psilocybin produces substantial and sustained decreases in depression and anxiety in patients with life-threatening cancer: a randomized double-blind trial. J Psychopharmacol. 2016;30:1181–97.
8. Ross S. Serotonergic hallucinogens and emerging targets for addiction pharmacotherapies. Psychiatr Clin North Am. 2012;35:357–74. doi:10.1016/j.psc.2012.04.002.
9. Johnson MW, Hendricks PS, Barrett FS, Griffiths RR. Classic psychedelics: an integrative review of epidemiology, therapeutics, mystical experience, and brain network function. Pharmacol Ther. 2019;197:83–102. doi:10.1016/j.pharmthera.2018.11.010.
10. Osmond H. A review of the clinical effects of psychotomimetic agents. Ann N Y Acad Sci. 1957;66:418–34. doi:10.1111/j.1749-6632.1957.tb40738.x.
11. Pahnke WN, Richards WA. Implications of LSD and experimental mysticism. J Relig Health. 1966;5:175–208. doi:10.1007/BF01532646.
12. Savage C, McCabe OL. Residential psychedelic (LSD) therapy for the narcotic addict. A controlled study. Arch Gen Psychiatry. 1973;28:808–14. doi:10.1001/archpsyc.1973.01750360040005.
13. Bedi G, Cecchi GA, Slezak DF, Carrillo F, Sigman M, De Wit H. A window into the intoxicated mind? Speech as an index of psychoactive drug effects. Neuropsychopharmacology.2014;39:2340–48. doi:10.1038/npp.2014.80.
14. Johnson M, Richards W, Griffiths R. Human hallucinogen research: guidelines for safety. J Psychopharmacol. 2008;22:603–20. doi:10.1177/0269881108093587.
15. Rice PL, Ezzy D. Qualitative research methods: a health focus. South Melbourne, VIC, Australia: OUP Australia; 2000.
16. Sandelowski M. Telling stories: narrative approaches in qualitative research. Image J Nurs Sch. 1991;23:161–66. doi:10.1111/j.1547-5069.1991.tb00662.x.
17. Atkinson P, Delamont S. Rescuing narrative from qualitative research. Narrat Inq. 2006;16:164–72. doi:10.1075/ni.16.1.21atk.
18. Cambria E, White B. Jumping NLP curves: a review of natural language processing research [review article]. IEEE Comput Intell Mag. 2014;9:48–57. doi:10.1109/MCI.2014.2307227.
19. Carrillo F, Mota N, Copelli M, Ribeiro S, Sigman M, Cecchi G, et al. Emotional intensity analysis in bipolar subjects [Internet]. arXiv [cs.AI]; 2016. Available from: http://arxiv.org/abs/1606.02231 [last accessed 4 Sep 2020].
20. Carson NJ, Mullin B, Sanchez MJ, Lu F, Yang K, Menezes M, Cook BL. Identification of suicidal behavior among psychiatrically hospitalized adolescents using natural language processing and machine learning of electronic health records. PLoS One. 2019;14:e0211116.
21. Carrillo F, Sigman M, Fernández Slezak D, Ashton P, Fitzgerald L, Stroud J, Nutt DJ, Carhart-Harris RL. Natural speech algorithm applied to baseline interview data can predict which patients will respond to psilocybin for treatment-resistant depression. J Affect Disord. 2018;230:84–86.
22. Coyle JR, Presti DE, Baggott MJ. Quantitative analysis of narrative reports of psychedelic drugs [Internet]. arXiv [q-bio.QM]; 2012. Available from: http://arxiv.org/abs/1206.0312 [last accessed 4 Sep 2020].
23. Zamberlan F, Sanz C, Martínez Vivot R, Pallavicini C, Erowid F, Erowid E, Tagliazucchi E. The varieties of the psychedelic experience: a preliminary study of the association between the reported subjective effects and the binding affinity profiles of substituted phenethylamines and tryptamines. Front Integr Neurosci [Internet]. 2018;12. doi:10.3389/fnint.2018.00054.
24. Garcia-Romeu A, Griffiths RR, Johnson MW. Psychedelic-associated addiction remission: an online survey. Drug Alcohol Depend. 2017;171:e69. doi:10.1016/j.drugalcdep.2016.08.199.
25. Garcia-Romeu A, Davis AK, Erowid F, Erowid E, Griffiths RR, Johnson MW. Cessation and reduction in alcohol consumption and misuse after psychedelic use.

J Psychopharmacol. 2019;33:1088–101. doi:10.1177/0269881119845793.

26. Johnson MW, Garcia-Romeu A, Johnson PS, Griffiths RR. An online survey of tobacco smoking cessation associated with naturalistic psychedelic use. J Psychopharmacol. 2017;31:841–50. doi:10.1177/0269881116684335.
27. Loper E, Bird S. NLTK: the natural language toolkit [Internet]. arXiv [cs.CL]; 2002. Available from: http://arxiv.org/abs/cs/0205028 [last accessed 4 Sep 2020].
28. Leskovec J, Rajaraman A, Ullman JD. Mining of massive datasets. 3rd ed. Cambridge, England: Cambridge University Press; 2020.
29. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python [Internet]. arXiv [cs.LG]; 2012. p. 2825–30.Available from: http://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf [last accessed 23 Feb 2021].
30. Deerwester S, Dumais ST, Furnas GW, Landauer TK, Harshman R. Indexing by latent semantic analysis. J Am Soc Inf Sci. 1990;41:391–407. doi:10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9.
31. Klema V, Laub A. The singular value decomposition: its computation and some applications. IEEE Trans Automat Contr. 1980;25:164–76. doi:10.1109/TAC.1980.1102314.
32. Field M, Mogg K, Mann B, Bennett GA, Bradley BP. Attentional biases in abstinent alcoholics and their association with craving. Psychol Addict Behav. 2013;27:71–80. doi:10.1037/a0029626.
33. Jajodia A, Earleywine M. Measuring alcohol expectancies with the implicit association test. Psychol Addict Behav. 2003;17:126–33. doi:10.1037/0893-164X.17.2.126.
34. Muller AC, Guido S. Introduction to machine learning with python: a guide for data scientists. Sebastopol (CA): O'Reilly Media; 2016.
35. Connor JP, Gullo MJ, White A, Kelly AB. Polysubstance use: diagnostic challenges, patterns of use and health. Curr Opin Psychiatry. 2014;27:269–75. doi:10.1097/YCO.0000000000000069.
36. John WS, Zhu H, Mannelli P, Subramaniam GA, Schwartz RP, McNeely J, Wu L-T. Prevalence and patterns of opioid misuse and opioid use disorder among primary care patients who use tobacco. Drug Alcohol Depend. 2019;194:468–75.
37. McCabe SE, West BT. The 3-year course of multiple substance use disorders in the United States: a national longitudinal study. J Clin Psychiatry. 2017;78:e537–44. doi:10.4088/JCP.16m10657.
38. Bogenschutz MP, Pommy JM. Therapeutic mechanisms of classic hallucinogens in the treatment of addictions: from indirect evidence to testable hypotheses: hallucinogens for addictions. Drug Test Anal. 2012;4:543–55. doi:10.1002/dta.1376.
39. Chwelos N, Blewett DB, Smith CM, Hoffer A. Use of d-lysergic acid diethylamide in the treatment of alcohol ism. Q J Stud Alcohol. 1959;20:577–90. doi:10.15288/qjsa.1959.20.577.
40. Grof S, Goodman LE, Richards WA, Kurland AA. LSD-assisted psychotherapy in patients with terminal cancer. Int Pharmacopsychiatry. 1973;8:129–44. doi:10.1159/000467984.
41. Kurland AA, Grof S, Pahnke WN, Goodman LE. Psychedelic drug assisted psychotherapy. In: Goldberg IK, Malitz S, Kutscher AH, editors. Patients with terminal cancer psychotheramacological agents for the terminally ill and bereaved. New York: Columbia University Press; 1973. p. 86–133.
42. Pisano VD, Putnam NP, Kramer HM, Franciotti KJ, Halpern JH, Holden SC. The association of psychedelic use and opioid use disorders among illicit users in the United States. J Psychopharmacol. 2017;31:606–13. doi:10.1177/0269881117691453.
43. Johnson MW, Garcia-Romeu A, Griffiths RR. Long-term follow-up of psilocybin-facilitated smoking cessation. Am J Drug Alcohol Abuse. 2017;43:55–60. doi:10.3109/00952990.2016.1170135.
44. Uthaug MV, Van Oorsouw K, Kuypers KPC, Van Boxtel M, Broers NJ, Mason NL, Toennes SW, Riba J, Ramaekers JG. Sub-acute and long-term effects of ayahuasca on affect and cognitive thinking style and their association with ego dissolution. Psychopharmacology (Berl). 2018;235:2979–89.