Supplementary methods

This supplement is a more detailed version of the Methods in the main body of the paper.

Recruiting platform and study eligibility

Data for this study were obtained as part of a larger online survey that included sets of questions from several different investigators and covered a wide range of topics. The subset of items analyzed here focused on social activity and changes in drug use and entrapment during the pandemic. Respondents were recruited and enrolled using the online crowdsourcing platform Amazon Mechanical Turk (mTurk). Within mTurk, Human Intelligence Tasks (HITs) are used to offer opportunities to participate in research activities or other tasks. A HIT describes the activity and the compensation amount, allowing potential respondents to choose whether to accept and complete the HIT. Eligibility requirements for the larger study were: residence in the United States (US), proficiency in English, age 18 years or older, and previous completion of at least 100 HITs (i.e., demonstrating acceptable performance on previous tasks). Additional inclusion criteria for our set of questions were self-reported use of one of the following classes of drugs during the six month period prior to screening: (1) alcohol only (not counting nicotine and caffeine); (2) prescribed or nonprescribed opioids; or (3) nonprescribed psychostimulants. For opioids and stimulants, use of any additional drug did not preclude enrollment. This was intended to be a nonclinical sample; we did not require problematic use for eligibility.

Data collection

Between September 3, 2020 and March 31, 2021, we sent HITs for our screening questionnaire to mTurk workers whose accounts were registered as belonging to adult US residents. A total of 13,608 respondents completed the 8-item screening survey and were
compensated $0.08. Among those who completed the screening, 5,112 (38%) were eligible for the full survey. Among these, 3,417 (67%) qualified based on alcohol use, and 1,695 (33%) qualified based on opioid or psychostimulant use. Those who completed the full survey, hosted on Qualtrics, were compensated $7.25.

To help ensure validity, 26 quality checks were programmed into the survey. Automatic unenrollment occurred if 3 or more quality checks were failed or if the survey was not completed within the allotted 4 hours. All responses were tracked by mTurk identification number and internet protocol (IP) address to avoid having multiple surveys completed by the same person. IP addresses were checked via IPHub to exclude use of virtual private networks or proxy addresses. The study was exempt from institutional review because no personally-identifiable information was collected, apart from IP addresses, which were deleted after verification of location in the US.

Among the 5,112 eligible screener respondents, 2,864 (56%) completed the full survey. We subsequently removed 249 (1%) of the completed surveys due to one of the following: (1) IP address (duplicate, outside the US, or indeterminant location); (2) improbably short survey completion time; or (3) inconsistent responses to drug-use items across the screening and full survey. The final number of surveys used for analysis was 2,615.

Eighty percent of respondents used only alcohol. Most respondents who reported use of opioids or stimulants (75% and 77%, respectively) also reported use of alcohol, and 16% of those who used either opioids or stimulants used both opioids and stimulants (see Venn diagram in "Supplement—Regression variables"). If a respondent provided data regarding changes in alcohol, opioid and/or psychostimulant use, these data were included in the analyses regardless of which drug class qualified them for the study during initial screening.
Measures of psychological characteristics potentially affected by the pandemic

Several brief psychometric instruments were incorporated into the survey, in addition to questions we developed regarding social conditions and changes in drug use. *Past-month loneliness* was assessed using the 20-item University of California, Los Angeles (UCLA) Loneliness Scale (Doryab et al., 2019; Russell, 1996); higher scores indicate greater loneliness (range: 20-80). *Past-month depression* was assessed using the Center for Epidemiologic Studies Short Depression Scale (CES-D-R-10) (Björgvinsson, et al., 2013; Miller, Anton, & Townson, 2008; Radloff, 1977), consisting of 10 items measuring past-week depression symptom severity across domains corresponding to DSM-5 criteria; higher values represent more depressive symptoms. *Past-month anxiety* was assessed using the Generalized Anxiety Disorder Scale (GAD-7; (Spitzer et al., 2006), a 7-item assessment of symptoms based on DSM-IV diagnostic criteria; higher values represent greater severity (range 0-21). *Entrapment* was assessed using two unidimensional scales (Gilbert & Allan, 1998): Internal Entrapment, based on 6 items (e.g., “I want to get away from myself,” “I feel trapped inside myself”) related to escape motivation stemming from internal stimuli (i.e., cognitions, emotions); and External Entrapment, based on 10 items related to escape motivation stemming of external conditions (e.g., “I feel trapped by other people,” “I feel powerless to change things”). In other contexts (such as prediction of newly emergent suicidal ideation in people with past suicide attempts), Internal and External Entrapment have differing associations (Höller et al., 2021), but in our sample, the two subscales were highly correlated.

Assessment of changes in drug use and entrapment

For each drug that was reported as being used in the past 12 months, the respondent was asked, "Has your use of _____ increased or decreased since the Covid-19 pandemic began?"
The possible answers were "Increased," "Decreased," "Not changed" or "I don't know."

Responses of "I don't know" were treated as missing data. Specific opioids listed on the survey included: "heroin," "fentanyl," "prescription opioids not prescribed to you," "opioids prescribed to you," "methadone not prescribed to you," "methadone prescribed to you," "Suboxone/Subutex (buprenorphine) not prescribed to you," and "Suboxone/Subutex (buprenorphine) prescribed to you." Specific psychostimulant items were: "powder cocaine," "crack/rock/freebase cocaine," "street meth (crystal meth, crank)," and "amphetamine pills (Ritalin, Adderall) not prescribed to you." There was a single item for change in use of "alcohol." For analysis, these responses were used to determine categorical outcomes for change in use of opioids, stimulants, and alcohol, as follows. Within each drug class, change in use was scored as "Increased" (if any item increased and none decreased), as "Decreased" (if any item decreased and none increased) or "Not changed" (if any item was not changed and none had increased or decreased, or if at least one item increased and at least one item decreased).

Our question about change in entrapment immediately followed the set of questions (described above) that assessed levels of internal and external entrapment. Specifically, change in entrapment was addressed by the question, "Related to these feelings of being trapped, how have these feelings changed since the Covid-19 pandemic began?" This question could be answered, "Much worse," "A little worse," "No change," "A little better," or "Much better."

Four types of socializing were assessed by asking how many of the past 30 days included the activity and "how satisfying" the activity was. In-person talking was described as "at least 30 minutes talking with other people in person," including "strangers, friends, family, co-corkers, or anyone else who comes to mind." In-person shared activities was described as "at least 30 minutes engaged in activities with other people in person," including "strangers, friends, family,
co-workers, or anyone else who comes to mind." *Online talking* was described as "at least 30 minutes communicating with other people online," including "strangers, friends, family, co-workers, or anyone else who comes to mind, that you engage with in a virtual space, such as texting, online message boards, forums, discussion groups, social media (e.g., Twitter, Facebook, etc.)." *Online shared activities* was described as "at least 30 minutes engaged in other activities (such as gaming) with other people online," including "strangers, friends, family, co-workers, or anyone else who comes to mind, that you engage with in a virtual space." Satisfaction from the specific type of socializing was assessed using a visual analog scale of 0-100, with a higher number indicating greater satisfaction. Self-described introversion was assessed with a visual analog scale (0-100) in response to the question, "I would consider myself an introvert."

**Data analysis**

*Initial models*

Statistical models were used to assess relationships between self-described introversion, social activity and changes in entrapment or drug use during the pandemic. Since some respondents did not engage in certain types of activity, and some activities were correlated (potentially multicollinear), a separate model was used for each of the 16 combination of one activity type (in-person talking, in-person shared activities, online talking or online shared activities) and one of the four outcomes (change in entrapment, change in alcohol use, change in opioid use or change in psychostimulant use). Mismatch between introversion and social circumstances was incorporated into each of the 16 social activity models as interaction effects between self-described introversion, the number of days that included the social activity, and the amount of satisfaction derived from the activity. Regressions were partially standardized (i.e., the regressors but not the outcomes were standardized) to facilitate interpretation and comparison of
regression coefficients and to help assess variable importance and effect size. Specifically, each regressor was standardized by subtracting the mean and dividing by the standard deviation based on the whole data set (as opposed to separately for each regression; see "Supplement—Regression variables" for raw means and standard deviations).

The results of these initial models indicated that, in general: (1) entrapment and drug use increased in respondents unsatisfied with their social activity, and (2) introversion and amount of activity had less influence than satisfaction, but (3) introversion interacted with satisfaction in certain models, with opioid use increased in unsatisfied extraverts, and psychostimulant use increased in unsatisfied introverts. Although the results of these models were informative, the piecemeal approach was not ideal because, by looking at each of the four types of social activity in isolation, these models did not adequately reflect the overall pattern of activity for each respondent. Adding or averaging across activity types would be similarly problematic because certain respondents did not engage in certain activities; we do not know whether they did so by choice or by force, so we do not know whether the activity would have been satisfying if it had occurred.

Hierarchical clustering

To address this shortcoming, we used hierarchical clustering (Maechler et al., 2018) to identify natural groups of respondents who shared similar patterns of activity, based on the frequency (days in last 30) and amount of satisfaction from each of the 4 kinds of activity. Based on the dendrogram from this clustering procedure (see "Dendrogram of hierarchical clustering results") and examination of the average levels of days and satisfaction within each cluster (as depicted in Fig. 2), we identified eight clusters that can be described based on the level of
satisfaction received from the type(s) of activity in which they engaged (see "Cluster table"). The chosen 8-cluster level cuts the dendrogram above the region where relatively short branches appear, thereby capturing the relatively large distinctions inherent in the data. The cluster numbers were assigned by the algorithm, and adjacent numbers would be expected to be more similar to each other than to clusters with relatively distant numbers.

**Dendrogram of hierarchical clustering results.** The dashed line represents the level chosen for analysis.

---

**Cluster table.** Summary of the most frequent activities (in-person vs. online) and the levels of satisfaction from these activities in each cluster, with a descriptive name based on this information. N indicates the number of respondents in the cluster.

| Cluster | Frequent type(s)       | Satisfaction | Brief description                  | N   |
|---------|-----------------------|--------------|------------------------------------|-----|
| 1       | Online only           | Satisfied    | Satisfied (online)                 | 217 |
| 2       | High both             | Satisfied    | Satisfied (both)                   | 189 |
| 3       | High in-person        | Satisfied    | Satisfied (in-person & text)       | 613 |
| 4       | In-person only        | Average      | Average (In-person)                | 481 |
| 5       | Low both              | Satisfied    | Satisfied (neither)                | 394 |
| 6       | Moderate to high both | Unsatisfied  | Unsatisfied (both)                 | 255 |
| 7       | Mostly in-person      | Very unsatisfied | Unsatisfied (in-person)          | 124 |
| 8       | Low both              | Unsatisfied  | Unsatisfied (neither)              | 342 |
Cluster heatmap. Level of satisfaction from each activity for each participant, grouped by cluster. The color indicates satisfaction, and the saturation (intensity) indicates the number of days, with more saturation indicating more days (i.e., the color is faded for respondents who rarely engaged in an activity, and invisible for those who did not engage at all). *Yellow indicates the activity was satisfying, and blue indicates the activity was unsatisfying.* Satisfaction and days were each standardized within the activity. The color scale is centered based on the range of standardized values (not on zero) because the distributions of values were not symmetric.

Satisfaction values were undefined (and therefore missing) when a respondent reported zero days for an activity; for these cases, clustering was conducted with satisfaction imputed as the overall sample mean for that variable. Subsequent assessment of the distinguishing features of the clusters (as in "Cluster table" and Fig. 2) was based on group means calculated only with nonmissing data (i.e., for a respondent with zero days of an activity, the respondent was omitted from calculation of the mean level of satisfaction).

Clusters 1 through 5 were generally satisfied with their activities; although the respondents in Cluster 4 had negative values (indicating below average satisfaction) from online
activities, they rarely engaged in such activities (i.e., "days" was also below average). Clusters 1 and 2 were the only ones with substantial days of online shared activities (for which gaming was the example given in the questionnaire); the difference between these two clusters is that Cluster 2 also had above-average levels of in-person activities, while Cluster 1 had below-average levels of in-person activities. Cluster 3 had frequent, satisfying in-person activities and online talking like Cluster 2, but without frequent gaming. Cluster 4 had slightly above-average levels of in-person activities and near-average satisfaction from these activities. Cluster 5 had some of the lowest levels of in-person activity, but received above-average satisfaction from it when it occurred. In contrast, Cluster 8 also had low levels of in-person activity, but found it unsatisfying. Clusters 6 and 7 had average or above-average levels of in-person activity, and were unsatisfied by it; they differed from each other in that Cluster 6 also had fairly frequent online activities (including gaming), which were also unsatisfying. For all types of activity, Cluster 8 had levels as low as or lower than any other clusters, and was unsatisfied. In short, Clusters 1 through 5 all engaged in various activities that were moderately to highly satisfying; Clusters 6 and 7 engaged in various activities but were unsatisfied by them; and Cluster 8 rarely engaged in any activity, and found it unsatisfying when they did engage.

Each cluster included a full range of introversion levels, and the average levels were similar, with credible intervals overlapping between all pairs of clusters except for Cluster 8, which had an interval that was higher than the intervals for Clusters 1 through 5, but not higher than 6 and 7 (i.e., higher than in the "satisfied" clusters, not higher than in the "unsatisfied" clusters.
**Introversion levels by cluster.** Medians and 90% credible intervals for level of introversion in each cluster

![Median introversion (with 90% CI)](image)

**Modeling procedures**

The cluster variable was used as a regressor in four models, one for each outcome variable (change in entrapment, change in alcohol use, change in opioid use, and change in stimulant use, respectively). The outcome was modeled as a function of cluster, introversion, and the demographic covariates described below. Another set of three models (one for each drug class) was used to assess change in drug use as a function of change in entrapment, along with self-described introversion and demographics as covariates; these models did not involve the social-activity variables and therefore did not involve the clusters. Demographics for the clusters are summarized in "Supplement–Demographics."
All models included four demographic covariates (age, sex/gender, race/ethnicity and Hispanicity). Race/ethnicity was included as a variable because it might reflect qualities related to self-identity and culture, which could be informative at the population level. For this variable, respondents could choose all that apply from a list ("African American/Black," "Asian," "Biracial," "Hispanic," "Native American, including Alaskan or Pacific Islander," "Indian," "Middle Eastern, "White/Caucasian," and "Other"). Responses typed in by respondents for the "Other" category were converted into appropriate responses from the remaining categories. Respondents who did not endorse any race/ethnicity were coded as "Unknown (Race not given)." Respondents who indicated more than one race/ethnicity (not including "Hispanic") or who only indicated "Biracial" were coded as "Multiracial." Categories with few respondents were combined or excluded. Specifically, "Indian," "Middle Eastern," and "Native American" had only 0-15 respondents for each outcome and were combined with the "race not given" category for all analyses. Hispanicity was treated as independent from the other race/ethnicity variables and comprised 65 respondents who endorsed "Hispanic" along with 2-4 other race/ethnicities, plus 133 who only endorsed "Hispanic"; those who endorsed other race/ethnicity items but not "Hispanic" were coded as "Nonhispanic," and those who did not endorse any race/ethnicity were coded as "Hispanicity unknown."

In all models, the outcome was treated as an ordinal variable in a Bayesian regression using weakly regularizing priors (Bürkner, 2017; Bürkner & Vuorre, 2019); specifically, the priors were a t-distribution with $\mu=0$ and $\sigma=2.5$ for thresholds, and a normal distribution with $\mu=0$ and $\sigma=3$ for all other coefficients. Posterior distributions from these regressions provided (1) credible intervals for the regression coefficients, (2) Bayesian p values [i.e., "probability of direction" for coefficients representing the probability that a parameter is nonzero (Makowski,
Ben-Shachar, & Lüdecke, 2019; Makowski, Ben-Shachar, Chen, et al., 2019)], and (3) model-fitted probabilities for each categorical outcome. Effects were evaluated based on the regression coefficients shown in the tables and forest plots; a coefficient with a credible interval that excludes zero provides evidence for an effect, and the distance from zero estimates the size of the effect. The relationships represented by these effects (especially those involving statistical interactions) are best understood by inspecting plots showing the probability of each ordinal level of the outcome as a function of the regressors; these plots, which also provide the most direct assessment of effect size (i.e., how much the outcomes are affected by the regressors) are provided in the main body of the paper or in the Supplementary material. Rather than intercept coefficients, the ordinal regressions have thresholds that are necessary for the regression but are not of direct interest. The most prevalent categories ("Cluster 3," "Female," "White," "Nonhispanic") were used as the reference categories for cluster, sex/gender, race/ethnicity, and Hispanicity, respectively. In the models of change in drug use as a function of change in entrapment, “No change” was the reference level for entrapment. In the tables and figures, point estimates represent medians, and error bands indicate 90% credible intervals. Credibly nonzero effects (Bayesian p < .1) are highlighted in the coefficient tables and forest plots to facilitate interpretation.

To assess the possibility of a secular trend in levels of social activity over the period in which we conducted the survey (such that earlier respondents might have differed from later respondents), we ran a sensitivity analysis modeling the effects of time as a sole regressor of the amount of each of the two in-person days variables (in-person talking and in-person shared activity). We also added time as a covariate to the model of each outcome as a function of cluster. In all models, the coefficient for time had a narrow credible interval centered near zero,
suggesting that levels of activity (and the effects of clusters derived from activity and satisfaction levels) did not vary substantially as a function of when the survey was taken.

References for supplementary material

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan [Bayesian inference; multilevel model; ordinal data; MCMC; Stan; R]. 2017, 80(1), 28 J Journal of Statistical Software. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C., & Vuorre, M. (2019). Ordinal Regression Models in Psychology: A Tutorial. 2(1), 77-101. https://doi.org/10.1177/2515245918823199

Doryab, A., Villalba, D. K., Chikersal, P., Dutcher, J. M., Tunminia, M., Liu, X., Cohen, S., Creswell, K., Mankoff, J., Creswell, J. D., & Dey, A. K. (2019). Identifying behavioral phenotypes of loneliness and social isolation with passive sensing: statistical analysis, data mining and machine learning of smartphone and Fitbit data. JMIR Mhealth Uhealth, 7(7), e13209. https://doi.org/10.2196/13209

Gilbert, P., & Allan, S. (1998). The role of defeat and entrapment (arrested flight) in depression: an exploration of an evolutionary view. Psychol Med, 28(3), 585-598. https://doi.org/10.1017/s0033291798006710

Höller, I., Rath, D., Teismann, T., Glaesmer, H., Lucht, L., Paashaus, L., Schönfelder, A., Juckel, G., & Forkmann, T. (2021). Defeat, entrapment, and suicidal ideation: Twelve-month trajectories. Suicide Life Threat Behav. https://doi.org/10.1111/sltb.12777

Maechler, M., P, R., A, S., M, H., & K, H. (2018). cluster: Cluster Analysis Basics and Extensions, R package version 2.1.2.

Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. Journal of Open Source Software, 4(40). https://doi.org/10.21105/joss.01541

Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., & Lüdecke, D. (2019). Indices of Effect Existence and Significance in the Bayesian Framework [Original Research]. 10(2767). https://doi.org/10.3389/fpsyg.2019.02767

Russel, D. W. (1996). UCLA Loneliness Scale (Version 3): reliability, validity, and factor structure. Journal of Personality Assessment, 66(1), 20-40.

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. Archives of internal medicine, 166(10), 1092-1097.