OSF REGISTRIES | Add New | My Registrations | Help | Donate

Observational Study of Cannabis

Embargoed registration | Updates

- Overview
- Metadata
- Files
- Resources
- Wiki
- Components 0
- Links 0
- Analytics
- Comments 0

Open practice resources

- Data
- Analytic code
- Materials
- Papers
- Supplements

Study Information

Hypotheses

We consider three primary hypotheses related to task outcomes associated with programming under the influence of cannabis: diversity, bugs, and speed. Finally, we hypothesize that these correlations will be less pronounced in heavy users.

We hypothesize that solutions to free-form programming problems by cannabis-using programmers exhibit greater diversity and creativity (see outcome measures below)

We hypothesize that programs written by cannabis-intoxicated programmers contain more bugs In particular, we hypothesize that this increase in bugs will primarily be in syntactic bugs (e.g., typos or misused variables) rather than semantics bugs due to evidence that acute cannabis use is more likely to impact short-term rather than long-term memory ([1], also see [5] for an older example)

We hypothesize that non-intoxicated programmers will take less time to complete the same coding tasks than intoxicated programmers (productivity)

Finally, we hypothesize that any observed relations between cannabis use and diversity, defect rates and speed (as above) will be less evident for users frequently self-reporting heavy THC-dominant cannabis use compared to more moderate users. Use frequency history will be calculated using the DFAQ-CU

Finally, we hypothesize that all of these effects will be less pronounced in participants with a history of heavy cannabis use than those with more moderate use histories.

Design Plan

Study type

Observational Study - Data is collected from study subjects that are not randomly assigned to a treatment. This includes surveys, “natural experiments,” and regression discontinuity designs.

Blinding

No blinding is involved in this study.

Personnel who analyze the data collected from the study are not aware of the treatment applied to any given group.

Is there any additional blinding in this study?

No - the motivation of this study is to understand the impact of cannabis use on programming in the context of job policies and productivity. Cannabis users typically know they have used cannabis before doing such activities.

Study design

We will recruit around 100 programmers who have previously used cannabis. Over zoom, we will observe these programmers program on various tasks, once while they intoxicated by cannabis (their own product, not research team provided) and once when they are not. While programming, we will record

[redacted]

Description

Show more

Registration type
OSF Preregistration

Date registered
April 19, 2023

Date created
April 19, 2023

Associated project
osf.io/akzvj

Category
Project

Publication DOI
No publication DOI

Subjects
Software Engineering
Computer Sciences
Physical Sciences and Mathematics

Affiliated institutions
This registration has no affiliated institutions

License

Tags
No tags

Citation
osf.io/g6fds

participant key strokes and response time and use these methods to answer our research questions. We will aim for around half of our participants to be regular cannabis users and the other half to be more moderate/non-regular users. Our goal is to have a large enough population for each group to permit a subpopulation analysis between those participants with a history of heavy cannabis use and those without. The document attached below contains more details on our experimental design.

- Observational Study of Cannabis Using Programmers.pdf

Randomization

We will randomize the order of the intoxicated and sober observational study participants for each participant.

Sampling Plan

Existing Data

Registration prior to analysis of the data

Explanation of existing data

Data is currently being collected when this pre-registration is submitted. The pre-screening form for our study is live. About half of data has been collected. None of the data has been analyzed. Hypotheses were generated before any data was collected but after ~5 pilot runs that are not included in the study data.

Data collection procedures

inclusion criteria: Programmers (such as graduate students or company employees) that are over the age of 21, have used cannabis at least once in the last year, have smoked or vaped cannabis, have a programming experience of at least the level of a senior undergraduate, and have familiarity with the programming language Python, be willing to attend two 1.5 hour remote sessions, one while intoxicated with cannabis, one while not. During these sessions, the participant will be required to participate from home or a place where they do not have to travel after consuming cannabis. They must also use a personal computer to avoid any tracking software on an employer-owned device.

exclusion criteria: Individuals who meet the inclusion criteria but are not willing to be observed during two recorded programming sessions. Individuals who have not used cannabis before or have not used cannabis in the last year or have not ever smoked or vaped cannabis are not eligible for this study.

I. Participants recruitment:
Participants will be recruited from the following sources:
a. Participants in previous cannabis and programming survey who wanted to be contacted for follow-up studies
b. Word of mouth through software contacts (snowball recruiting)
c. Social media posts on Twitter and applicable subreddits or public email lists
d. Posters/flyers that we put in public locations
A pre-screening survey will be used to see if potential participants meet our criteria. This prescreening survey is included in this application. In all of the forms of recruitment, there will be a link to the pre-screening survey. The prescreening survey will be electronic, hosted on qualtrics.
Our goal is to recruit 100 participants in total.

I. Before the experiment:
Some participants who meet the study inclusion criteria will be asked to complete an additional survey that contains more detailed information about their programming background, cannabis usage history, and previous experiences with cannabis and programming, and study session availability. These participants will be chosen on a primarily first-come first serve purpose. However, as one of our hypotheses relates to the impact of cannabis usage history (e.g., do programmers who are heavy cannabis users see the same disruption as those who are light users), we may prioritize participants based on this so that we have a large enough subsample for this comparison. This survey will be hosted on qualtrics, and participants will be contacted through the email provided in the prescreening survey. These potential participants will

be selected from the pre-screening responses primarily on a first-come-first-serve basis. However, we aim in this study to have fairly even groups of both moderate and heavy cannabis users so that we can compare the effects between the two. Therefore, we may recruit some participants out of order so that this is the case. This initial survey after the prescreening will contain an electronic consent form, though we note we will also get verbal consent at the beginning of each zoom session. It will also include a short (under 5 minute) training video for the remote sessions. Should participant information from this second survey be inconsistent with that from the first (e.g, they are not actually eligible to participate, or it is unclear if they are), they will be removed from the study. In this survey, participants will indicate when would be a good time to observe them programming when intoxicated with cannabis and when would be a good time to observe them while programming sober.

III. Experimental procedure: Session Scheduling: There will be two remote sessions over Zoom that we ask each participant to join. On one, we will ask to observe the participant program at a time they are intoxicated with cannabis. This will be an observation of acute cannabis use: participants will be currently under the influence while participating and not experiencing lingering effects from past use. The other will be performed when they are not currently under the influence of cannabis. For the intoxicated session, we ask the participants to choose their own time (e.g., a time they would normally use cannabis), send us pictures of the product, and reschedule if they have not used cannabis before the session. While we do not want to ask participants specifics about their work hours, participants will be encouraged to schedule at a time outside of work when they don't have other commitments. Additionally, participants will need to be in a place where they don't need to travel afterwards for at least two hours for the using session. They will also be asked to use a personal computer. We will not ask participants specific location information (e.g., home address). However, we will ask participants to verbally confirm they are in a safe location where they don't need to travel at the beginning of the session, and we will also mention this request in the scheduling confirmation email.

We will not give participants specific instructions on the amount of cannabis to consume. We will only ask that they do not schedule their session at a time they anticipate using cannabis edibles as this ingestion method can lead to very different acute impacts than smoking or vaping. Due to the ingestion method, we will ask that participants consume cannabis within 5-10 minutes before the session starts as smoked cannabis can impact the system fairly quickly. Should the participant be unable to or forget to take a picture of the product, we will continue the session, but remove this participant from the main statistical analysis and consider them as a qualitative use case where we have less certain information. The only constraint on scheduling time other than researcher availability is that the order of the two sessions will be counterbalanced for participants - that is, we will ask half to schedule their intoxicated session first and the other half to schedule their sober session first. Participants are asked to share their videos, audio, and screen. And we will take screen recordings of their session to aid in the analysis. In particular, these recordings serve three purposes: 1) They allow us to capture data features such as mouse movement and window movement that we are not able to capture with Qualtrics and the survey platform 2) they allow us to note verbal aspects of programming (e.g., some programmers may talk through problems under their breath) and 3) they are another form of data redundancy.

Participants will be compensated $80 for completing both sessions and $20 for completing one.

No files selected

Sample size

50-100 people

Sample size rationale

Due to the novelty of the work, the expected effect size is a bit uncertain. However, for some preliminary power analyses, we use effect sizes observed of acute cannabis use on various cognitive processes also used in programming (see [6]). We would want a sample size large enough to expect to observe differences in some of these areas (e.g., that could be observed in the cognitive baseline tests we give participants).

Here is the estimated sample size we would need to expect observing differences for each of these cognitive areas assuming alpha = 0.01 (lower than 0.05 to preemptively consider multiple comparison correction in this estimate) and statistical power = 0.80 (note: these are likely lower bounds as we will have less control over the dosage and the like compared to lab studies):
Attention: observed effect is g = 0.223 -> need 239 participants (unlikely to observe this with our sample sizes below)
Executive functions: observed effect is g = 0.370 -> need 89 participants
Speed of processing: observed effect is g = 0.384 -> need 83 participants
Verbal learning: observed effect is g = 0.688 -> need 28 participants
Verbal Memory: observed effect is g = 0.513 -> need 48 participants
Working Memory: observed effect is g = 0.514 - need 48 participants

While observing attention differences is likely beyond the scope of this study, we otherwise favor targeting about 100 participants.

Stopping rule

No response

Variables

Manipulated variables

While an observational study, participants come in twice -> once sober and once 5-10 minutes after vaping or smoking cannabis in a manner that they normally do

No files selected

Measured variables

Measured variables in the session:

a. A set of small programming tasks (e.g., change this program code to remove a bug, provide missing code for this partial program to obtain required functionality, etc.) (15-25 min)
b. Three open-ended "interview style" coding questions that require substantial creativity(40-50 min)
c. A short/small number of standardized psychological tests (e.g., stroop interference task and a spatial reasoning task, etc.) that have previously been found to be impacted by acute cannabis use (i.e. active cannabis intoxication) as an independent measure of cannabis intoxication (5-10 min)
d. Short and open-ended questions about current cannabis intoxication level and their experience in the study (2-3 min)
For a, c, and d, the participants will answer the questions on Qualtrics which is approved for hosting sensitive human subjects data. For b, we will observe participants program in Codespaces, an online coding platform that is private (requires research team authentication to enter) and run through GitHub. Data saved on Codespaces will only include keystrokes and button press timing - we will not collect any identifying or sensitive information on code spaces. We've devised the following tools to aid our data collection:
1. Local History, a VSCode extension that is used to auto-save the participant-generated files and send them to our server.
2. Keyboard Macro, a VSCode extension that is used to record the user-input keystrokes with timestamps.
3. A bash script that intercepts the build process.
The three tools help us record the information that we can use to analyze the coding behavior of the programmers.
The participants are asked to complete a post-survey to reflect on their performance.

Outcome Measures

To test our hypothesis we will use the following metrics:
Creativity: For free-response questions, we will measure the diversity of programming solutions between groups (in terms of algorithms chosen, including algorithmic problem solving style, asymptotic complexity, etc.).
Additionally, we will adapt existing metrics for creativity in programming to our

study (see https://ieeexplore.ieee.org/abstract/document/6901522 and https://dl.acm.org/doi/abs/10.1145/3051457.3054003 for example creativity metrics programming).
Bug error rates: We will use a combination of held-out test suites and manual inspection to determine the correctness of a given programming solution and the number of defects that it contains.
Productivity: We will have a battery of programming tasks (more than could be completed in both sessions). Productivity will be measured by the number of tasks the participant completes (or the time taken to complete each task) and how this number varies between the sessions.
Cannabis-usage history as a moderating variable: We will use portions of the DFAQ-CU to measure self-reported cannabis usage history.

We also collect basic demographics (age, gender, weight) and programming history (e.g., years of professional programming experience)

No files selected

Indices

No response

No files selected

Analysis Plan

Statistical models

Before we do any analysis, we will run summary statistics (average score on held-out tests, average typing speed, etc) for all participants. We will also check for learning effects on both the qualtrics and the leet code problems. We will report the results of these checks in our paper. Now, here are the hypothesis we are pre-registering.

Hypothesis 1: We hypothesize that programs written by cannabis-intoxicated programmers contain more bugs.

This will be done by running a per-problem t-test between high-solutions and sober solutions to each leet code problem. The t-test will be done on the solution scores for each problem. The t-test type will be determined after a test for normality. We will also run a t-test for the qualtrics problems. We will do both a general "High-vs sober" test again per problem, as well as a paired test on problems that participants did each time. The null hypothesis will be rejected if high programmers have lower scores on the majority of leet-code problems when high.

Hypothesis 2: We hypothesize that programs written by cannabis-intoxicated programmers will be written slower and run slower.

For this one, we will run three tests:
1) Typing speed. We will compare the average characters typed per minute between high participants and sober participants during the leet code problems using a t-test. We expect it to be lower for high participants.
2) We will look only at those participants who got a correct leet code or Qualtrics coding solution (For leet-code, they must have gotten at least 75% of the held-out tests). We will then, per problem, use a t-test to compare how long it took to get to that solution.
3) We will check how long the held-out tests took to run for those participants that got all problems correct. We expect the held out tests to take longer to run for high participants.

Hypothesis 3: We hypothesize that solutions to free-form programming problems by cannabis-using programmers exhibit greater diversity and creativity.

To measure this, we will be using a two-phase approach:

1) We will be qualitatively analyzing free-form participant responses to the leet code problems. We will group them by the method. These solution methods are started by looking at the most popular solutions posted to leet code.
2) We will then run a chi-squared test across participants to see if there is a

different ratio of methods in the sober vs high solutions.
3) We will reject the null hypothesis if a) there is a different distribution of methods between the two groups and b) the high group contains more methods and/or the distribution is more even across methods.

Hypothesis 4: We hypothesize the impact of cannabis use while programming will be less for those that are heavy cannabis users vs those that are moderate users. We will divide are participants into three groups: Light cannabis users (those who use cannabis at most 3-4 times a month), Moderate cannabis users - those who use 1-2 times per week) and heavy users (those who use more than 2 times per week. We will then compare the light users to the heavy users on the tests for Hypothesis 1 and 2 above. We expect to see a larger difference between the light users then the heavy users. Additionally, we will see if within participant differences in typing speed and qualtrics correctness + time are correlated to cannabis usage level.

No files selected

Transformations

There are three places we will need to code qualitative data for this experiment's hypothesis.

1) Qualtrics coding output problems
2) Qualtrics short coding problems
3) Leet code problems

For the first one, it will be graded by one expert annotator using the following scale:
1) Correct
2) Correct other than new lines or programming-language-specific output (e.g., [])
3) Incorrect

For the second one, it will be graded on a similar scale:
1) Correct
2) Correct other than syntax errors (algorithmically incorrect)
3) Incorrect - Complete
4) Incorrect - incomplete

For the leet code problems, they will be categorized into solution methods. These methods will initially be based on the methods of the most popular solution submitted to leet code. However, we will use an iterative qualitative approach using the actual data to build the final rubric. When building this rubric, the researchers will not know if a solution was produced while high or sober. Solutions will be categorized and the rubric modified until saturation is reached. Two researchers will categorize each problem by method and then disagreements will be discussed to come to a consensus if possible. Beyond problem category, leet code problems will also be graded for style: We will record if 1) They used comments, 2) They made more tests, 3) They used asserts, 4) They used understandable variable names, 5) They made helper functions, and 6) If they had duplicated code.

Beyond qualitative analysis, each leet-code problem will be given a correctness score. This will be done by running held-out tests for each problem. We will record the correctness and the time for each solution (so we can capture solution complexity). Each test bench will be run three times with a different seed such that the time we get is more accurate.

For quantifying cannabis use history as needed for our third hypothesis, we will use the scoring associated with the DFAQ-CU (the validated cannabis usage history assessment we are using).

Inference criteria

$P < 0.05$, we will use BH correction for multiple comparison correction within a hypothesis as listed above.

We will be using two-tailed tests for all analyses

Data exclusion

A participant must complete both coding sessions to be included in the analysis. They must be sober for one and high for the other. If this is not the case, their data will be discarded.

Leetcode problems: If a participant never passes above 25% of the test cases on any leet-code problem, their leet-code data will be discarded as they do not have enough Python knowledge to participate.

Qualtrics Problems - Stroop Test: If a participant indicates that they are color blind and/or gets gets more than half of the Stroop test problems wrong, they will be excluded from the analysis of this question.

Qualtrics Problems - Paper Folding: If a participant rushes through this in under 3 seconds per question (less than 30 seconds total), they will be excluded. Participant data may also be excluded if there are technical difficulties indicated on the experimenter log for this participant (e.g., browser crashes, internet goes down).

Qualtrics Problems - Coding: If a participant rushes through in under 3 seconds per question, they will be excluded from that problem. Participant data may also be excluded for a problem if there are technical difficulties indicated on the experimenter log or video recording for this participant (e.g., browser crashes, internet goes down, double clicks, and accidentally skips a problem).

Missing data

Participants with data missing for a particular statistical test (e.g., keystroke data missing for a programming session) will not be included in that statistical test. Missing data will not be generated or interpolated or filled with an "average response".

Exploratory analysis

Any tests beyond those in the box above will be labeled exploratory.

The things we plan to look at in this manner are:

1) How program correctness evolves over time between sober and not sober
2) Stroop effect problems -> We have a tentative hypothesis that those in the high group may actually perform better on this than those in the sober group based on some preliminary work, but this is very tentative.
3) The impact of demographic features such as Gender, Age, Weight, etc...
4) The impact of the amount of cannabis used (e.g., those that report very high vs those that report less high) -> though we strongly suspect those that report very high will do worse than those who report less high.
Other things that emerge during data analysis.

Other

Other

No response