PNAS
www.pnas.org

Supplementary Information for:

Cannabis Increases Susceptibility to False Memory

Authors: Lilian Kloft, Henry Otgaar, Arjan Blokland, Lauren A. Monds, Stefan W. Toennes, Elizabeth Loftus, Johannes G. Ramaekers

Corresponding authors:

Lilian Kloft (l.kloft@maastrichtuniversity.nl), Johannes G. Ramaekers (j.ramaekers@maastrichtuniversity.nl)

This PDF file includes:

Supplementary Information Text

Participants

Sample size was determined by use of a power calculation based on results of the misinformation method obtained in a previous experiment. In a field study on the effects of alcohol on memory and susceptibility to suggestive cues (1), sober and intoxicated participants (n = 67) were asked to commit a mock crime. Intoxicated participants displayed an increased tendency to go along with misleading questions compared to sober participants with a medium to large effect size (η^2 = .12, which equals Cohen's f = .37). An a priori statistical power analysis by means of G*Power 3.1 (2) showed that to detect comparable effects using a slightly more conservative estimate (Cohen's f = 0.30), a sample size of 64 participants would be required, using a repeated measures between-subjects ANOVA with a power of 0.80, and a significance level alpha of 0.05.

The majority of participants completed testing at the lab site in Maastricht (n = 56). Participants were recruited via online as well as offline advertisements posted around Maastricht University and Sydney University. Subjects were screened using a medical history and drug use questionnaire, and underwent a general medical examination including routine laboratory tests through a medical supervisor. Inclusion criteria were: occasional cannabis use (minimum 1/month and maximum 2/week on average during the past 12 months); aged between 18 and 40 years; free from psychotropic medication; good physical health as determined by medical examination and laboratory analysis; absence of any major medical, endocrine and neurological condition; body mass index (weight/height^2) between 18 and 28 kg/m^2; good knowledge and understanding of the English language (≥5 years of English language education), and written informed consent). Exclusion criteria were: history of drug abuse (other than cannabis) or addiction (determined by the medical questionnaire, drug questionnaire and medical examination); pregnancy or lactation

(determined by pregnancy test on test day); hypertension (diastolic >90 mmHg; systolic >140 mmHg); history of psychiatric disorders; liver dysfunction; (serious) side effects due to previous cannabis consumption and history of cardiac dysfunctions (arrhythmia, ischemic heart disease, etc.).

Design, Doses and Administration

The study was conducted according to a double-blind, placebo-controlled, 2 (Group: Treatment vs. Control) by 2 (Time: Time 1 vs. Time 2) mixed design with Group as a between subjects factor and Time as a within-subjects factor. All participants were randomly assigned to one out of 4 possible randomization sequences, counterbalancing the order of the treatment and VR scenario (see Fig. S1). Treatment consisted of Bedrobinol, which is medicinal cannabis containing 13.5% THC and <1% cannabidiol (CBD). Placebo consisted of Knaster Hemp, which is a non-psychoactive herbal plant mixture containing 0.0% THC. Bedrobinol was forwarded by the Office for Medicinal Cannabis (Ministerie van Volksgezondheid, n.d.) at the site in Maastricht, and by the company Novachem at the Sydney site.

On separate test days, each subject inhaled the vapor of a single dose of Bedrobinol (300 µg THC/kg bodyweight) or a placebo (150 mg Knaster Hemp). This dose of cannabis has been used in previous studies and can be seen as an average dose (3). Knaster Hemp has been used as placebo in previous studies (4). Administration took place by using the Volcano vaporizer (volume 8l). To prepare the vaporizer, its temperature was set at position nine (225 °C). Ten minutes before administration, the vaporizer was switched on to heat up. By placing the valve balloon on the filling chamber (containing either Bedrobinol or Knaster Hemp), hot air was blended with the THC or Knaster Hemp. For administration, participants were instructed to inhale deeply and subsequently hold their breath for 10 s before exhaling. This continued until the balloon was

emptied. An investigator was present in the room while administration took place. Treatment preparation was done by a different researcher from the one performing the administration and testing.

Measures and Materials

Deese/Roediger-McDermott (DRM) paradigm. The DRM was used to induce spontaneous false memories. Two parallel versions of the DRM were used (day 1: version A, day 2: version B), which were counterbalanced with treatment order. For each version, the study phase consisted of 15 DRM word lists containing ten words each (total 150 stimuli; first ten words of the respective lists by Roediger et al., 5). Normative data have shown that these lists vary in both their backward associative strength (BAS, index of the associative strength between the list items and the critical item) and their inter-item associative strength (see 5, 6). An overview of the lists by version and their BAS is displayed in Table S2. The two versions did not statistically significantly differ in BAS [t (28) = .41, p = .68]. Lists were presented visually via PowerPoint, starting with an announcement of the list number (e.g., List 1), followed by the respective study words being shown one-by-one in the center of the screen (duration 2 s per word). All stimuli were separated by a 2 s interstimulus interval, during which the plus symbol was shown in the center of the screen. The total duration of the study phase was 640 s. Participants were instructed to pay attention and try to remember the words as their memory for these words would be tested later in the session.

For each DRM version, there were two testing phases: one administered immediately (approximately ten minutes after end of study phase), and one administered 7 (± 1) days later. These will be referred to as the immediate and the delayed tests, respectively. Thus, two test versions were created per DRM version, resulting in total in four test instances per participant. The immediate version consisted of 75 words: 45 previously presented words (words 1, 3, and 5 from

each list), 10 new words critically related to the studied lists (*critical lures*), 10 new words related to the studied lists (*related lures*, partly taken from words 11-15 from the original DRM lists, and partly from https://wordassociations.net/en), and 10 new unrelated words (*unrelated words*, adopted from other, non-presented DRM lists). The delayed version consisted of 100 words: 55 presented words (10 of these had been already presented at immediate test), 15 critical lures (10 from immediate test), 15 related lures (5 from immediate test), and 15 unrelated words (5 from immediate test). Before testing commenced, participants were instructed to indicate whether they recognized the words from the previous list presentation (yes or no). The words appeared on the computer screen one at a time in random order. The study and immediate testing phases were separated by a subjective high measurement and two 5 min filler tasks (attention tasks: Psychovigilance Test and Deary-Liewald reaction time task; 7, 8).

Outcome measures included *true memory rates* (the proportion of studied words correctly recognized at test), false alarm rates for *critical lures* (the proportion of critical lures, i.e., new, strongly related words, that are incorrectly recognized at test, a measure of false memory), false alarm rates for *related lures* (proportion of incorrect recognition of new, related words) false alarm rates for *unrelated words* (proportion of incorrect recognition of new, unrelated words), and net accuracy (ratio of true memory to all memory, an indication of overall ability to discriminate between studied and unstudied items).

Misinformation paradigm. In order to investigate suggestion-based false memory formation, an adjusted version of the misinformation paradigm was used (9, 10). On separate test days, participants were involved in two distinct crime scenarios, simulated in a fully immersive virtual environment. The virtual reality headset *HTC* Vive was used. The device uses “room scale” tracking technology in order to turn the environment into a 3D space in which the user can move

freely. Motion-tracked controllers were used so that the participant could interact with the environment. VR has been previously applied successfully in eyewitness memory studies conducted by our lab (11). An image section of both VR scenarios is displayed in Fig. S2, and respective videos can be viewed on the Open Science Framework (https://osf.io/k5v8c/).

Interviews to assess true and false memory were conducted about 30 min post the VR simulation (i.e., *immediately*) and once during the follow-up session (7 ± 1 days later, i.e., *delayed*). Before the interview, subjects were instructed to answer with yes or no, to be as truthful as possible, and to guess if they did not know the answer. Interviews consisted of non-leading questions about truly presented details (e.g., “Were the seats on the train blue?”), leading questions about suggested details (e.g., “It was a black purse, right?”), and non-leading questions about non-presented details (e.g., “Was there a cat in the bar?”). Details of the latter category varied in their event plausibility (i.e., included questions about plausible details, such as person selling snacks on train, but also implausible details, such as clown on the platform). For the eyewitness scenario, the immediate interview consisted of 25 questions (15 presented, 5 suggested, 5 non-suggested), and the delayed of 29 questions (15 presented, 9 suggested, 5 non-suggested; 20 new and 9 old items). For the perpetrator scenario, the immediate interview contained 25 questions (15 presented, 5 suggested, 5 non-suggested), and the delayed of 27 questions (15 presented, 7 suggested, 5 non-suggested; 20 new and 7 old items[1]). The order of the questions remained the same for all participants. A Qualtrics file on a tablet was used to record the answers.

Eyewitness scenario. In the eyewitness scenario participants were passive witness to the physical attack on a police man and a security guard by one man (the attacker). In this scenario the

[1] For analysis, 2 questions about presented details from the perpetrator scenario were excluded due to VR-related difficulties

crime took place on a platform at a train station and participants witnessed the scenario from inside the train, among other virtual passengers. Prior to the simulation participants were instructed to imagine that they were on a train traveling with a friend who was sitting opposite to the participant in the train, to remain seated during the simulation, and that at one point a crime would take place, which they should observe. Two min after the crime occurred, the friend (co-witness) engaged in a monologue directed towards the participant and recalled some aspects of the attack. She provided correct information (e.g., the attacker kicked the security guard) as well as misinformation (e.g., there was a police dog on the platform). The simulation ended after she provided all information.

Perpetrator scenario. The perpetrator VR-scenario was designed in a way that participants were in a bar setting where they were able to walk and explore the bar. Furthermore, some avatars would engage in a monologue when participants approached them. Prior to the simulation participants received instructions about the scenario. They were instructed to imagine themselves in the role of a student who had lost their job and was in urgent need of money to pay their rent, thus deciding to obtain some money from someone in their local bar. They were encouraged to explore the bar and instructed to grab the strap of the purse when the people who were playing a game started cheering. The trigger stimulus was presented after 2 min. A motion-tracked controller was used to simulate the purse that was visible for the participants in the VR simulation. A leather handle was mounted onto the controller in order to create the haptic feel of a purse strap. The scenario was manually ended when participants grabbed the controller.

Physiological measures. Blood samples (5 ml) were taken at THC peak concentration and after completion of the testing procedure (~5 and ~120 min post-administration). All blood samples were centrifuged and the serum was frozen at -20°C and transported to an external lab for pharmacokinetic assessments of cannabinoids (analyses described in 12).

Measures of heart rate were taken shortly before (2 min baseline) and during the VR simulations, using the Garmin watch Forerunner® 15 heart rate monitoring belt (recorded every 20 s).

Subjective high. In line with other work from our lab (see e.g., 13), participants were asked to rate how affected they were by cannabis at the present moment by placing a vertical dash on two visual analogue scales (100 mm), stating their subjective feeling of cannabis influence (*subjective high*, ranging from "totally not under the influence of cannabis" to "very much under the influence of cannabis") and in comparison to previous experiences with cannabis (*subjective experience*, "much less under the influence" to "much stronger than usual"). Subjects rated this ~20 min, ~80 min and ~110 min after administration.

Procedure

A full timetable of procedures can be viewed in Table S3. Participants who passed all screening procedures were invited for a training session in order to get acquainted with the virtual reality (VR) program, the vaporizer used for administration of the cannabis/placebo, and other tests used during test days. Testing consisted of two similar test days, which were scheduled 7 (± 1) days apart, and one follow-up meeting 7 (± 1) days after the second test day. All meetings took place in a laboratory at the study site, with the exception of 6 final follow-up meetings that were conducted via phone/email due to the participant's unforeseen unavailability. Participants were requested to abstain from drug use 7 days and from alcohol use 24 h prior to testing, to have a light breakfast/lunch before, to not consume any caffeine-containing products throughout the day, and to arrive well-rested. Drug and alcohol screens were conducted before the start of every testing day (incl. the follow-up session). An additional urine pregnancy test was performed for women. All breath alcohol concentration readings showed 0.00 (missing data $n = 7$). All pregnancy tests

showed a negative result. In case of a positive drug test before the start of the first testing day, participants were rescheduled for a later date. In case of a positive drug test before the start of the second testing day or final follow-up session, a blood sample was taken but the test day was carried out nonetheless. The sample was analyzed for active THC metabolites, and data from participants with THC levels >2.0 ng/ml was later excluded from further analysis (n = 1).

Following the cannabis/placebo administration and the first blood sample, all memory and other cognitive procedures were conducted at fixed intervals during ~120 min post-administration. Upon completion of all tests, a final blood sample was taken. The participant was obliged to remain at the lab until minimum 3 h after drug administration. The researcher would determine, with help of a discharge form, whether it was deemed safe for the participant to go home. After study completion, participants received a short debriefing explaining the goals of the study.

Statistical approach

To test drug effects on DRM memory recognition performance in the immediate condition, a 2 (Drug: cannabis vs. placebo) x 4 (Level of association: old words, critical lures, related lures, unrelated words) repeated measures analysis of variance (ANOVA) was conducted. The same was repeated for delayed DRM performance. To compare the groups' eyewitness memory performance at immediate and delayed test respectively, we conducted two separate 2 (Group: cannabis vs. placebo) x 3 (Question type: true, suggestive, irrelevant) repeated measures ANOVA. Equivalent analyses were conducted for perpetrator memory. When a statistically significant interaction effect was detected, simple main effects were assessed. If a statistically significant main effect was detected, post-hoc comparisons were conducted using the Bonferroni correction. T-tests were used for pairwise comparisons. Visual inspection of mean scores was also used to aid interpretation. A difference was considered statistically significant for p-values < 0.05. Cohen's d (pairwise

comparisons) and ω^2 (ANOVA) were calculated as effect size estimates. All values are reported including two decimals, except for p-values where three decimals are reported. The assumptions underlying all analyses were checked. For ANOVA, assumptions were checked by visual inspections of boxplots for normality. No gross violations of assumptions were detected for ANOVA. When sphericity was violated, a Greenhouse–Geisser correction was applied to the degrees of freedom.

To exclude any potential differences caused by the study site, all main analyses to investigate effects of cannabis in the two false memory paradigms were repeated, excluding the participants recruited in Sydney (n = 8, *Site* section, SI). Only results where a change in effects was found are reported. Subjective high results were analysed using repeated measures ANOVA. All analyses were conducted using JASP, version 0.11.1 (14).

Additional analyses

Site. All effects detected in the DRM immediate analyses remained unaffected when including only the Maastricht participants (n = 56). Similarly, the interaction between Drug and Level of association was statistically significant in this reduced sample; however, statistical significance was not maintained for some effects according to the simple main effects analysis (n = 54). Whereas statistically significant effects had been detected before true memory and false memory of critical lures, the p-values for these now exceeded the alpha level of .05 (p = .075 and p = .12, respectively). These effects thus appear not very robust.

With regard to the immediate analyses of the eyewitness misinformation task, the interaction remained statistically significant, but the simple main effect for questions about non-suggested details lost significance (p = .11; n = 28). For the immediate analyses of the perpetrator

scenario, the main effect of Group did not attain statistical significance (p = .06). All other effects reported remained unchanged.

Subjective high and THC serum concentration. Mean values for subjective high and subjective experience for both placebo and cannabis conditions as a function of time are shown in Fig. S3. Repeated measures ANOVA revealed a statistically significant difference in subjective high [F (1, 62) = 529.50, p < .001, ω^2 = .81] and subjective experience [F (1, 61) = 374.55, p < .001, ω^2 = .77] between both conditions. A summary of mean (*SD*) and range of THC, THC-OH and THC- COOH concentrations in serum as a function of time after smoking is given in Table S4.

Tables

Table S1. Subject demographics and drug history

Native language (#)	
English	10
Dutch	13
German	18
Other language	23
Level of education[a] (#)	
High school	30
Bachelor's degree	30
Master's degree	3
Other	1
Drug history [*M* (*SD*)]	
Age of first use	17.8 (2.5)
Years since using cannabis	4.6 (2.2)
Frequency/month	3.1 (1.9)

Notes.

[a] Level of education was measured in terms of highest level of education completed.

Table S2. DRM lists with backward associative strength (BAS) parameters

Version A		Version B	
List = Critical Lure	Mean BAS	List = Critical Lure	Mean BAS
Anger	0.181	High	0.109
Black	0.094	Lion	0.125
Bread	0.179	Man	0.131
Car	0.348	Mountain	0.157
Chair	0.284	Music	0.210
City	0.180	Needle	0.205
Cold	0.315	Pen	0.176
Cup	0.154	River	0.145
Doctor	0.234	Shirt	0.242
Foot	0.184	Sleep	0.452
Fruit	0.288	Smell	0.294
Girl	0.129	Soft	0.191
King	0.240	Sweet	0.223
Rough	0.165	Trash	0.118
Smoke	0.197	Window	0.221
Mean	*0.212*	*Mean*	*0.200*
SD	*0.072*	*SD*	*0.087*

Notes.

All mean BAS values have been calculated based on Roediger et al. (2001).

Table S3. Overview of testing procedures

	Procedure	Time after treatment (minutes)
Testday 1		
	Drug and alcohol screens	0
	Baseline questionnaires	0
	Administration cannabis/placebo	0
	Blood sample 1	5
	DRM A study phase	10
	Subjective high 1	20
	Cognitive tasks	22
	DRM A immediate test	32
	Virtual Reality scenario	37
	Cognitive tasks/Questionnaires	50
	Misinformation interview	75
	Subjective high 2	80
	Cognitive tasks	82
	Subjective high 3	110
	Blood sample 2	115
Testday 2		
	Drug and alcohol screens	0
	Baseline questionnaires	0
	Follow-up testday 1:	0
	DRM A delayed test	0
	Misinformation delayed interview	0
	Administration cannabis/placebo	0
	Remaining procedures of testday 2 are equal to testday 1	
Follow-up		
	Drug and alcohol screens	-
	Baseline questionnaires	-
	Follow-up testday 2:	-
	DRM B delayed test	-
	Misinformation delayed interview	-
	Debriefing	-

Notes.

Schedule is hypothetical and deviation could occur.

Table S4. Serum concentration values

	Mean	SD	Range
Sample 1[1]			
THC	77.31	64.37	0.90 - 325.70
THC-OH	4.37	3.15	0.50 - 15.90
THC-COOH	10.02	7.58	0.00 - 39.30
Sample 2[2]			
THC	4.19	2.16	0.70 - 10.90
THC-OH	1.66	0.75	0.60 - 4.20
THC-COOH	11.21	7.30	1.70 - 44.80

Notes.

[1]Sample taken immediately after administration. Missing data $n = 3$

[2] Sample taken after last testing procedure. Missing data $n = 10$

Figures

<table>
<tr><th>Participants (N=64)</th><th colspan="2">Week 1</th><th colspan="2">Week 2: Testday 1</th><th colspan="2">Week 3: Testday 2</th><th>Week 4</th></tr>
<tr><td>Sequence A (n=16)</td><td rowspan="4">Medical exam</td><td rowspan="4">Training Day</td><td>Drug + Eyewitness</td><td rowspan="4">Immediate memory tests</td><td rowspan="4">Follow-up memory tests (Testday 1)</td><td>Placebo + Perpetrator</td><td rowspan="4">Immediate memory tests</td><td rowspan="4">Follow-up memory tests (Testday 2)</td></tr>
<tr><td>Sequence B (n=16)</td><td>Drug + Perpetrator</td><td>Placebo + Eyewitness</td></tr>
<tr><td>Sequence C (n=16)</td><td>Placebo + Eyewitness</td><td>Drug + Perpetrator</td></tr>
<tr><td>Sequence D (n=16)</td><td>Placebo + Perpetrator</td><td>Drug + Eyewitness</td></tr>
</table>

Fig. S1. Schematic representation of counterbalanced randomization sequences A-D with the variables treatment (drug vs. placebo) and mock crime scenario (eyewitness vs. perpetrator)

Fig. S2. Screenshots from eyewitness (left) and perpetrator (right) virtual reality scenarios

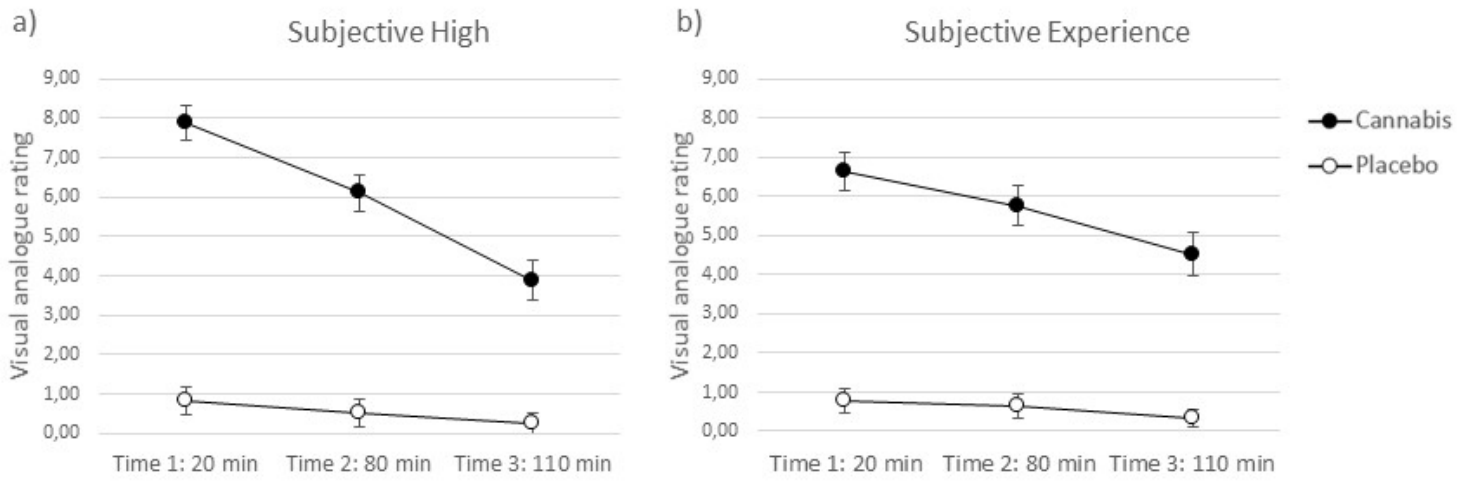

Fig. S3. Visual analogue scale ratings (0-10 cm) of subjective high (a) and subjective experience (b) as a function of time. Error bars represent 95% CIs.

SI References

1. K. Van Oorsouw, H. Merckelbach, T. Smeets, Alcohol intoxication impairs memory and increases suggestibility for a mock crime: A field study. *Applied Cognitive Psychology* **29**, 493-501 (2015).
2. F. Faul, E. Erdfelder, A.-G. Lang, A. Buchner, G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods* **39**, 175-191 (2007).
3. E. Theunissen *et al.*, Rivastigmine but not vardenafil reverses cannabis-induced impairment of verbal memory in healthy humans. *Psychopharmacology* **232**, 343-353 (2015).
4. J. Ramaekers *et al.*, Cannabis and tolerance: acute drug impairment as a function of cannabis use history. *Scientific Reports* **6**, 26843 (2016).
5. M. A. Stadler, H. L. Roediger, K. B. McDermott, Norms for word lists that create false memories. *Memory & Cognition* **27**, 494-500 (1999).
6. H. L. Roediger, J. M. Watson, K. B. McDermott, D. A. Gallo, Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review* **8**, 385-407 (2001).
7. S. Loh, N. Lamond, J. Dorrian, G. Roach, D. Dawson, The validity of psychomotor vigilance tasks of less than 10-minute duration. *Behavior Research Methods, Instruments, & Computers* **36**, 339-346 (2004).
8. I. J. Deary, D. Liewald, J. Nissan, A free, easy-to-use, computer-based simple and four-choice reaction time programme: the Deary-Liewald reaction time task. *Behavior Research Methods* **43**, 258-268 (2011).
9. E. F. Loftus, Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory* **12**, 361-366 (2005).
10. H. Otgaar, M. L. Howe, N. Brackmann, T. Smeets, The malleability of developmental trends in neutral and negative memory illusions. *Journal of Experimental Psychology: General* **145**, 31-55 (2016).
11. T. Romeo, H. Otgaar, T. Smeets, S. Landstrom, D. Boerboom, The impact of lying about a traumatic virtual reality experience on memory. *Memory & Cognition* **47**, 485-495 (2019).
12. S. W. Toennes, S. Hanisch, W. Pogoda, C. Wunder, A. Paulke, Pitfall in cannabinoid analysis—detection of a previously unrecognized interfering compound in human serum. *Analytical and Bioanalytical Chemistry* **407**, 463-470 (2015).
13. J. Ramaekers, G. Kauert, E. Theunissen, S. W. Toennes, M. Moeller, Neurocognitive performance during acute THC intoxication in heavy and occasional cannabis users. *Journal of Psychopharmacology* **23**, 266-277 (2009).
14. JASP Team (2019) JASP (Version 0.11.1). p [Computer software].