Between persistence and flexibility: the neuromodulation of cognitive control

Prochazkova, L.

Appendices

Appendix A

Chapters’ Author contributions

Chapter 2

Based on: Prochazkova, L., & Hommel, B. (2020). Altered states of consciousness and creativity. In *Creativity and the Wandering Mind* (pp. 121-158). Academic Press

Author contribution statement: L.P. and B.H. developed the theoretical formalism of the topic. L.P. wrote the first draft of the article under the supervision of B.H. B.H. provided critical revisions and supervised the project.

Chapter 3

Based on: Prochazkova L, Colzato L, Hommel B,. Metacontrol of event-file management: More selective handling after Focused-Attention than after Open-Monitoring Meditation (in preparation).

Author contribution statement: L.S.C. had the original idea to investigate the effect of meditation on the Event-file task. All three authors, together with Dr. Roberta Sellaro (R.S.), conceptualized the study's design and approved the present methodology of the experiment. L.P. collected and analysed the data under the supervision of R.S. L.P. interpreted the data and wrote the first draft of the article under the supervision of B.H. B.H. and L.S.C. provided critical revisions to the first draft.

Chapter 4

Based on: Prochazkova, L., Lippelt, D. P., Colzato, L. S., Kuchar, M., Sjoerds, Z., & Hommel, B. (2018). Exploring the effect of microdosing psychedelics on creativity in an open-label natural setting. *Psychopharmacology*, *235*(12), 3401-3413.

Author contribution statement: The first two authors contributed equally. L.P. had the original idea to investigate the effect of microdosing on creativity. L.P. and D.P.L. conceptualized the study's design and approved the present version of the experiment. Both L.P. and D.P.L. collected and pre-processed the data and the final analyses were carried out by D.P.L. L.P. and D.P.L. wrote the first draft of the article under the supervision of B.H. who provided critical revisions with the other co-authors.

Chapter 5

Based on: Prochazkova, L., van Elk, M., Marschall, J. C., Rifkin, B. D., Fejer, G., Schoen, N., Fiacchino, D., Fejer, G., & Hommel, B. Microdosing psychedelics and its effect on creativity: Lessons learned from three double-blind placebo controlled longitudinal. Psychopharmacology (submitted).

Author contribution statement: LP had the original idea to investigate effects of microdosing on creativity. The design of the study was developed by L.P., J. M., B.H. & M.v.E. The study was set up by L.P. and run together with J.M., B.D.R., N.S., D.F., G.F. Moreover, L.P. analyzed the data and wrote the first draft of the article under the supervision of B.H. and M.v.E. who also contributed to data interpretation. All authors contributed to the manuscript reviews and approved the final version of the manuscript.

Chapter 6.

Based on: Prochazkova, L., van den Broek, K., van Elk, M., Marschall, J. C., Rifkin, B. D. & Hommel, B. The effect of microdosing psychedelics on mood and cognition: Null effects in two randomized, double-blind, placebo-controlled longitudinal trials (in preparation).

Author contribution statement: The design of the study was developed by L.P., J. M., B.H. & M.v.E. The study was set up by L.P. and run together with J.M., B.D.R. L.P. and B.v.K. together analyzed the data and wrote the first draft of the article under the supervision of B.H. All authors contributed to the manuscript reviews and approved the final version of the manuscript.

Chapter 7:

Based on: Kočárová, R., Procházková, L., Rifkin, B., Rak, J., Hommel, B., Horáček, J., Carlos Bouso, J. (under review*)* The effects of use of secretion from the glands of Incilius alvarius containing 5-MeO-DMT on mental health, psychological flexibility and life-satisfaction: an observational prospective study.*Nauture (submitted).*

Author contribution statement: The first two authors contributed equally. R.K. and L.P. conceptualize the study, selected the methodology and wrote the original draft of the paper. L.P., J.R., and B.R analyzed the data. B.H, J.H. and J.C.B. contributed to the supervision of the project and manuscript reviews. All authors approved the final version of the manuscript.

Chapter 8:

Based on: Prochazkova L, Speer P.H., Zhang C., Trutti A., Smidts A., Boksem A.S., Hommel Metacontrol can account for interindividual variability in the impulsivity-compulsivity spectrum (in preparation).

Author contribution statement: L.P. conceptualize the study and wrote the original draft of the paper. The experimental data were collected by Speer P.S., under the supervision of A.S. and A.S.B. Note that this study, which served other theoretical purposes, included the collection of additional data and measurements that are not reported here. Also note that our study/analytical design and our hypotheses were developed after the data collection was completed but before we had access to the data. The authors are indebted to Speer, Smidts, and Boksem for providing us access to the data and to contribute to our analyses. All authors contributed to the data interpretation and manuscript reviews. All authors approved the final version of the manuscript.

Appendix B

Supplementary Material for Chapter 3

Supplementary Materials Experiment 1

Experiment 1A

S1. *Power calculation Experiment 1*

Sample size of at least (N=40 per group) was chosen based on Power calculation. A power calculation ran with G*power version 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009) indicated that at least a total of 80 participants should be included within the analyses in order to detect a significant interaction with a moderate effect size ($\eta p^2 = .06$), given alpha =0.05 and power = 0.80.

S3.1. *Data-screening*

Trials with missing responses R1 were excluded. Data for R1 were not further analyzed and trials with a missing R2 (0.1% of all trials with a correct R1) were also excluded from analysis.For the RTs and ERs analysis, we excluded anticipation responses (<150 ms) and missing responses (> 2000ms). For the RTs analyses the incorrect responses to S1 and S2 were further excluded. Additionally, participants with an average RT deviating more than 3 standard deviations from the sample's mean were removed from all analyses as well as those with very high error rates (around chance level 50%). Finally, Participants with history of psychiatric illness, previous meditation experience or reporting recent drug/medication use were excluded from the final analyses.

S2. *Participants excluded*

Participants

In order to attain optimal power of the analyses nighty-four Leiden University students were initially recruited via online university system and offered course credit or a financial reward of 6.50 euros for participating in the current study. Once recruited, all participants were screened for demographics, psychiatric illness and drug use. Six participants were excluded due to very high error rates (around chance level 50%) and one because of very slow reaction times (more than two standard deviation from the sample mean), three participants had to be excluded due to history of psychiatric illness, two participants failed to follow experimental instructions (one used a telephone during the task, other had a Caps Lock turned on) and 2 participants were excluded due to software error (data were not saved).

3.1.3. Apparatus and stimuli

Audio priming stimuli

Each of the priming session lasted approximately 17 minutes. In the FAM condition, a male voice guides participant to sustain their attention on their own breathing, to observe the quality of that attention and revert their attention back to their breathing whenever the mind starts to wander, in order to gradually lead the participant in to a single focus. In the OMM condition, the same male voice guides participants to pay attention to the present moment and to simply notice their own awareness, feelings, thoughts, and bodily sensations without holding on to them and emotionally reacting to them. The focus should be expanded to non-jugging awareness of ''here and now''.

Affect Grid

The Affect Grid (Russell, et al., 1989), is a single-item scale suited for repeated assessment of people's subjective affective states. The scale consists of a 9 × 9 grid, in which the horizontal axis constitutes affective valence (ranging from unpleasantness to pleasantness), and the vertical axis depicts perceived arousal (ranging from high arousal to sleepiness). Two scores

can be derived from the scale, one for pleasure and one for arousal. The Affect Grid has been shown to have good reliability, convergent validity, and discriminant validity.

Event-file task

The E-Prime 2.0 software system (Psychology Software Tools, Inc., Pittsburgh, PA) was used to generate the task and collect all the responses. The task, originally developed by Hommel (1998), was adapted from Colzato et al. (2012c; 2013). In this task two stimuli and two responses are presented (Figure 2). Each trial begins with presentation of a response cue for the first response (R1) and the response has to be carried out only after a first stimulus is presented (S1). The stimuli consisted of yellow or green coloured images of a banana or of an apple and the manual response (R1) should be made after the onset of the S1. The second stimuli presented (S2) is composed from the same, partly the same or completely different perceptual feature than the previous stimuli (S1). The S2 stimuli require a manual binary choice response (R2) to the shape of the second stimuli (S2) (an apple or banana). The task measures binding-related effects by examining partial-repetition costs related to combinations of stimulus features (shape and colour in our case) and combinations of stimulus features and the response. The correct R1 was signaled in advance of S1 (through a left- or right-pointing arrowhead), so that S1 and R1 could be varied independently, which was necessary to create orthogonal repetitions and alternations of stimulus shape and response. In addition, the colour feature was also varied by presenting the apple or banana in green or yellow (see Colzato et al., 2013).The task comprised of practice block of 32 trials before the meditation, and additional 10 practice trials after meditation. The practice followed by experimental block of 192 trials, presented in a random order. Experimental trials were equally distributed across 8 conditions, resulting for the combinations of stimulus features (shape and colour) and responses, which could all either repeat or alternate, thus creating a 2 × 2 × 2-factorial design.

Experiment 1B

S4.1. ***Data-screening***

Trials with missing responses R1 were excluded. Data for R1 were not further analyzed. Trials with a missing R2 (0.08% of all trials with a correct R1) were also excluded from analysis. For the RTs and ERs analysis, we excluded anticipation responses (<150 ms) and missing responses (> 1500 ms). For the RTs analyses the incorrect responses to S1 and S2 were further excluded. Additionally, participants with an average RT deviating more than 3 standard deviations from the sample's mean were removed from all analyses as well as those with very high error rates (around chance level 50%).

S4.2. ***Participants***

Fifty-one Leiden University students were initially recruited via online university system and offered course credit or a financial reward of 6.50 euros for participating in the current study. Once recruited, all participants were screened for demographics, psychiatric illness and drug use. Nine participants were excluded due to very high error rates (around chance level 50%) and one because of very slow reaction times (more than two standard deviation from the sample mean), one participant had to be excluded due to extensive meditation experience prior to the experiment.

Experiment 2

S5.1. *Data-screening*

Trials with missing responses R1 were excluded. Data for R1 were not further analyzed. Trials with a missing R2 (0.07% of all trials with a correct R1) were also excluded from analysis. For the RTs and ERs analysis, we excluded anticipation responses (<150 ms) and late or missing responses (> 1500ms). For the RTs analyses the incorrect responses to S1 and S2 were further excluded. Additionally, participants with an average RT deviating more than 3 standard deviations from the sample's mean were removed from all analyses as well as those with very high error rates (around chance level 50%).

S5.2. *Participants*

Fifty-seven Leiden University students, were recruited via online university system and were rewarded by course credit or a financial compensation of 6.50 euros per each session for participating. Seven participants were excluded as they failed to attend second testing session. Next, six participants were further excluded due to very high error rates (around chance level 50%) in the first session and two were excluded due to very high error rates in the second session. After participants were screened for demographics, two participants had to be excluded – one due to history of psychiatric illness and one for extensive meditation experience.

S5.3. *The Analyses Event file*

The task comprised of identical procedure as in the first experiment. Experimental trials were again equally distributed across 8 conditions, resulting for the combinations of stimulus features (shape and colour) and responses, which could all either repeat or alternate, thus creating a 2 × 2 × 2 - factorial design. The effect of meditation on the updating of stimulus–response episodes was assessed by submitting R2 correct reaction times (RTs), and percentage of errors (PEs) to separate 2 × 2 × 2 × 2 × 2 factorial ANOVAs with 2 levels for Group (OMM vs. FAM) and 2 levels for Response, 2 levels for Shape, and 2 levels for Color corresponding to the repetition vs. alternation, respectively entered as within-subject factors and 2 levels for session (1 session vs. session 2) were entered as the between subject factor to control for possible order effect.

Appendix C

Supplementary Material for Chapter 4

S1. Truffle alkaloids content analyses

Methanol and formic acid were of LC-MS grade and were purchased from Sigma-Aldrich (Czech Republic). Ultra pure water, 18.2 MΩ-cm, was produced by a Smart2Pure 12 water purification system (Thermo Scientific, Germany). The stock solutions of psilocin and psilocybin were prepared in methanol and the stock solutions of norbaeocystin and baeocystin in 50%(*v/v*) methanol (at a concentration 1 mg/mL and stored at -20 °C. A working solution contained a mixture of analytes and was diluted with methanol to 10 µg/mL.

Sample preparation

Capsules were opened and 100 mg of inner solid was weighted to dark glass tubes with screw caps. The samples were dissolved in 5 mL of methanol pre-purged with nitrogen gas. A rack with samples was subsequently covered with aluminum foil and vortexed for 150 min. The mixtures were centrifuged for 15 min at 2,000 rpm (25°C). 100 µL aliquot was diluted with 900 µL of 0.1% (*v/v*) formic acid. Samples were 100× diluted with 0.1% (*v/v*) formic acid because of high concentrations of psilocin and psilocybin. 5µL was injected into the LC-MS and samples were prepared at duplicate.

LC conditions

The LC system used was an Agilent 1290 Infinity (Agilent Technologies, USA) and conditions were as follows: a column Phenomenex Kinetex F5, 2.1 × 100 mm, 1.7 µm 100 Å, with pre-column, gradient elution with 0.1% (*v/v*) formic acid (mobile phase A) and methanol containing 0.1% (*v/v*) formic acid (mobile phase B) with a flow rate of 250 µL/min. The injection volume was 5µL with 3 s needle wash in a flush port. The gradient setup was: 0-5 min from 95% A to 20% A, 5-5.5 min from 20% A to 0%, 5.5-6.5 min back to 95% A and equilibration at the same level to 10 min. The valve arrangement: 0.5-6 min to the MS source.

MS conditions

The tandem MS used was 6460 Triple Quad LC/MS (Agilent Technologies, USA) with Jet Stream Electrospray Ionisation Source. A dynamic multiple reaction monitoring (dMRM) method was used. In brief the conditions of instrument were as follows: positive ion mode, ionisation voltage of 2300 V, a source temperature of 340°C and gas flow rate10 L/min. The values for sheath gas were 400°C and 12 L/min, and for the nebulizer 30 psi. Data was acquired and evaluated with MassHunter software (Agilent Technologies, USA).

Alkaloid detectability and concentrations

The evaluation of the developed analytical method (Hajkova et al. 2016) encompassed the determination of Limit of Detection (LOD) and Limit of Quantitation (LOQ), and of the applicable concentration range for each compound studied. The LOD indicates the amount of an analyte that is minimally detectable without guarantees regarding the bias or measurement errors in the results by an assay (c.q. not quantified). LOQ, on the other hand,

does indicate the minimum amount of an analyte that is quantifiable. For the here described analyses LOD was determined as 3 times the ratio of signal to noise and LOQ as 10 times the ratio of signal to noise. LODs and LOQs for the alkaloids in the analyzed samples are shown in Table S1.

Table S1 contains the alkaloid concentrations for each sample type separately. The concentrations were very similar across samples for all alkaloids and the differences between estimates were lower than the estimated measurement errors for the results obtained when the data is averaged across samples. As such, in the Results we reported the concentrations for the four alkaloids collapsed across the three samples.

Supplementary Table S1 LOD was calculated as 3×S/N (signal to noise ratio) and LOQ as 10×S/N (Note: The left part of the table shows the final concentrations measured owing to LC-MS; the right side of the table shows the calculation on 1 g of capsule inner solid).

Table S2 contains the alkaloid concentrations for each sample type separately. The concentrations were very similar across samples for all alkaloids and the differences between estimates were lower than the estimated measurement errors for the results obtained when the data is averaged across samples. As such, in the Results we reported the concentrations for the four alkaloids collapsed across the three samples.

Table S1 LOD and LOQ values

	ng/ml		*µg/g (ppm)*	
Alkaloid	**LOD**	**LOQ**	**LOD**	**LOQ**
Psilocybin	0.5	1.0	0.3	0.5
Psilocin	1.0	5.0	0.5	2.5
Norbaeocystin	5.0	10.0	2.5	5.0
Baeocystin	1.0	5.0	0.5	2.5

Note: Left part of the table shows final concentrations measured owing to LC-MS; right side of table shows calculation on 1 g of capsule inner solid); LOD = Limit of Detection; LOQ = Limit of Quantitation.

Table S2 Results of observed analytes for each sample type separately (µg/g [ppm])

Alkaloid	**0.22g sample**	**0.33g sample**	**0.44g sample**
Psilocybin	1557	1595	1632
Psilocin	76	87	93
Norbaeocystin	9	8	8
Baeocystin	30	31	33

Appendix D

Supplementary Material for Chapter 5

S1. Truffle alkaloids content analyses

Methanol and formic acid were of LC-MS grade and were purchased from Sigma-Aldrich (Czech Republic). Ultra pure water, 18.2 MΩ-cm, was produced by a Smart2Pure 12 water purification system (Thermo Scientific, Germany). The stock solutions of psilocin and psilocybin were prepared in methanol and the stock solutions of norbaeocystin and baeocystin in 50%(*v/v*) methanol (at a concentration 1 mg/mL and stored at -20 °C. A working solution contained a mixture of analytes and was diluted with methanol to 10 µg/mL.

	ng/ml		*µg/g (ppm)*	
	LOD	**LOQ**	**LOD**	**LOQ**
norbaeocystin	5.0	10.0	2.5	5.0
baeocystin	1.0	5.0	0.5	2.5
psilocybin	0.5	1.0	0.3	0.5
psilocin	1.0	5.0	0.5	2.5

Dried at 25°C, vacuum	µg/g (ppm)
norbaeocystine	< 2.5
baeocystine	74.5±3.7
psilocybine	1976.0±45.7
psilocin	110.1±4.0
recalculation of PSLC to PSLCB	153.2±6.1
sum	2129.2±85.2

S2. Inter-ratter agreement AUT

Table S2. Shows inter-rater agreement in the ratings of creativity scores the three experiments for the AUT and the PCT-d task. Cronbach's alpha was used as a measure of internal consistency.

Cronbach's Alpha

	Experiment 1			
	fluency	flexibility	originality	elaboration
Baseline (towel)	0.897	0.891	0.89	0.74
Session 1 (pen)	0.979	0.936	0.896	0.749
Session 1 (brick)	0.976	0.898	0.911	0.697
Session 2 (bottle)	0.953	0.95	0.898	0.765
Session 2 (newspaper)	0.965	0.959	0.945	0.648
	Experiment 2			
	fluency	flexibility	originality	elaboration
Baseline (brick)	0.966	0.954	0.818	0.661
Baseline (towel)	0.981	0.841	0.781	0.672
Session 1 (bottle)	0.854	0.810	0.682	0.699
Session 1 (newspaper)	0.981	0.726	0.721	0.701
Session 2 (pen)	0.985	0.804	0.709	0.631
Session 2 (mop)	0.995	0.798	0.743	0.698
	Experiment 3			
	Convergent	fluency	originality	Flexibility/elaboration
Session 1 (PCT-d)	0.948	0.798	0.749	NA
Session 2 (PCT-d)	0.952	0.740	0.816	NA

S3. Sample size calculation - Power analyses

Experiment 1 and Experiment 2

The power calculation for the acute AUT scores (administered 3 times) indicated that a total of 78 participants (with two equal-sized groups of n = 39) should be included within the analyses in order to detect a significant interaction with a moderate effect size (Effect size *f* = 0.25), given alpha = 0.05 and power = 0.80. A power calculation for the PCT scores (administered 2 times) indicated that a total of 82 participants (with two equal-sized groups

of n = 41) should be included within the analyses in order to detect a significant interaction with a medium effect size (Cohen's *f* = 0.25), given alpha = 0.05 and power = 0.80.

Experiment 3

A power calculation for the PCT-d scores (administered 2 times) in cross-over within subject design indicated that a total of 24 participants should be included within the analyses in order to detect differences between conditions with a moderate effect size (ηp^2 = .06), given alpha =0.05 and power = 0.80.

Mega-analyses

A power calculation for hierarchical linear regression (with 6 covariates) indicated that a total of 82 participants should be included within the analyses in order to detect differences between conditions with a moderate effect size (Partial R^2 = .09), given alpha =0.05 and power = 0.80.

S4. Supplement Experiment 1

Tasks and Questionnaires

Alternate Uses Task (AUT). Divergent creativity was assessed by means of the AUT (Guilford 1967) in Experiment 1 and 2. The participants were asked to report as many creative uses of an everyday object. Participants had 5 minutes per item to think of as many possible uses for the object (e.g., towel, pen, brick, mop or bottle). Different items were used for each session, but the order of items presented across time was consistent. By keeping the order consistent we eliminated possible co-founding factors of item difficulty (e.g., availability heuristics) to interact with the result. Conversely, the same pair of AUT items were presented to all participants at each testing sessions and thus the test battery remain equally challenging for both condition across time.

As usual, the AUT reports were rated by two different raters according to five different variables. *Fluency* = the total number of responses; *Flexibility* = the number of different categories of responses; *Elaboration* = the degree to which the person elaborates on their response (each 'elaboration' receives one point, so that the response "using a brick to prevent a door from slamming shut (1), when it is windy (2)" earns two elaboration points); *Originality* = the uniqueness of a response in that sample (Responses that have also been mentioned by 1% or less of the other participants receive 2 points for originality, responses that have been mentioned by 1-5% of the participants receive 1 point); and the ratio between originality and fluency (originality/fluency) as an additional index of divergent thinking that was previously

suggested as a more parsimonious measure of divergent quality (Hocevar & Michael, 1979.; Runco & Albert, 1985) and was shown to be affected by psychedelics in previous research (Kuypers et al., 2016; Mason et al., 2019).

Picture Concept Task (PCT). The (PCT) (Wechsler, 2003, Hurks et al., 2010) is a visual creativity task commonly used as a measure of convergent thinking. In this task, 17 stimuli consisting of a matrix of pictures (either 2x3, 3x3 or 3x4) were presented. For each stimuli participants had to find single common association between one picture from each row by clicking on the items and writing down the common connection (please see Supplementary materials, Section S3.1). In line with previous studies participants had 30 seconds per item to find the solution (Mason et al., 2021). Because the task assumes there is only one correct solution to each trial, the task has been previously compared to the RAT assessing convergent thinking (Hurks et al., 2010). This is because, to complete the task, one should converge on the correct solution while inhibiting inappropriate or less obvious associations or incorrect solutions. The final PCT scores were calculated as the sum of all correct responses.

PCT example: Example trial from the Picture Concept Task. The subject has to identify the common association between 1 item from each row. Left image shows the correct convergent solution (“water” by connecting picture 3, 5, 7).

Subjective effects. Subjective drug effects were measured, utilizing several tailored questions. Specifically, participants were asked about how psychoactive they found the microdoses to be (*i.e., How intensively do you perceive microdosing effect at the present moment?*). The intensity of their experience was ranked on a Likert-type scale with ratings ranging from zero to one hundred where zero referred to "no effects", and one hundred referred to "extremely strong psychedelic effects". Secondly, participants reported their guesses regarding their condition allocation at every test session (i.e., *Do you think you are using active psilocybin doses or a placebo?*). Participants were asked to select one from three possible responses: i) active psilocybin, ii) not sure iii) placebo.

Sign-ups and exclusion criteria

Prior to the study participants were asked for contra-indications, including a prior diagnosis or family problems with schizophrenia, psychosis, mania, or borderline. We also excluded participants who indicated to have an addiction, who had serious physical health issues (e.g., diabetes or brain injury) and who lacked proficient English language skills (as the study was conducted in English). All participants provided their written informed consent to participate.

A total of 103 participants signed up for the psychedelic workshop by means of social media and MI & PSN contact details and 80 participants filled in the initial screening forms. Only 77 healthy participants free from contra-indication, including a prior diagnosis or family problems with schizophrenia, psychosis, mania, or borderline disorder, were invited to attend the microdosing event and baseline measurement at the Leiden University. The first MI & PSN workshop took place in Leiden and the related study was advertised by word-of-mouth at Leiden University. As such the majority of participants were Leiden University students. All participants provided informed consent before starting the study procedure, which was approved by the local ethics committee (Leiden University, Institute for Psychological Research). We provided participants with the guidelines as specified by the researchers involved in this project, asking them to comply with the proposed microdosing schedule, self-administering the dose max. 1.5 h prior to the lab session, to remain blind to their condition, to self-administer. The duration of 1.5 h after ingesting the microdosing capsule is based on the observation that the subjective effects of a higher dose of psilocybin and the plasma levels tend to peak around 90 min after intake (Passie et al. 2002).

Subjective effects Results

Interestingly, participants in the active psilocybin condition did not differ from the placebo condition regarding their estimates of group allocation at the first (χ^2 (3, N = 59) = 0.587, p = 0.746, *Cramer's* V = 0.1) and the second follow-up session (χ^2 (3, N = 59) = 6.21, p = 0.1, *Cramer's* V = 0.324). The majority of participants reported being uncertain about their group allocation and showed a similar distribution of false positive and negative estimates (see Table 4: Experiment 1). More specifically, 26,6% of participants in the placebo condition believed to be in the active condition at the end of the two-week dosing period. Notably, two participants who were allocated in the control group, experienced very strong placebo effects, and requested smaller dosage from PSN during the trial. This further suggests presence of salient placebo effects as well as the success of the blinding manipulation.

Next, the group differences in strength of the perceived psychoactive effects were analyzed. The *rm*ANOVA indicated a main effect of session (F (2,110) = 4.499, p = 0.013, $\eta 2_p$= 0.08)

suggesting that the subjective microdosing strength changed over time while controlling for experimental condition. The follow up t-test indicated that the diminishing microdosing effects were observed between the initial workshop (*M* = 23.77, *SD* =21.82) and the last experimental session (*M* = 15.68, *SD* = 16.86); *t*(56) = 2.91, *p* = 0.005), for both placebo and active microdosing condition. Nevertheless, the main effect of group (*F* (1,55)= 0.97, *p* = 0.75, $\eta 2_p$= 0.002*)* and the interaction between group and session (*F* (2,110) =0.94, *p* = 0.39, $\eta 2_p$= 0.017*)* were not-significant, suggesting that participants in active psilocybin condition did not differ in ratings of psychedelic strength from placebo. As can be seen in Table 5. (Experiment 1) participants rated the strength of the psychedelic around 20% (100% referred to extremely strong psychedelic effects) and the psychedelic ratings diminished over time. The result again confirmed that the blinding manipulation was successful and the presence of placebo effects.

Previous experience and subjective effects

We further assessed the effect of previous psychedelic experience on subjective drug ratings. Interestingly, independent sample t-tests (with Welsh correction) indicated that participants, who had previous psychedelic experience rated the microdosing drug effect stronger at the acute 2 session, (*M* = 19.87, *SD* = 10.56) compared to participants who did not have prior psychedelic experience, (*M* = 8.81, *SD* = 10.56), *t* (49.74)= 2.69, *p* = 0.01, *d* =0.713), regardless of their actual microdosing condition. No effect of previous psychedelic experience was found at baseline and first acute session (*Ts* ≤ 1.13, *Ps* ≥ 0.28*)*. Follow-up analyses indicated that this effect was mainly driven by participants in the placebo condition (see Figure 3). Specifically, participants who had previous psychedelic experience and were in the placebo condition reported significantly higher placebo effects in terms of microdosing strength, (*M* = 22.6, *SD* = 21.5) in the acute 2 session than those who had no prior experience with psychedelics, (*M* = 6.9, *SD* = 8.9), *t*(21.44)= 2.6, *p* = 0.016, *d* =0.951). No significant differences for ratings of psychedelic strength were found in the active condition with regards to previous psychedelic experience (*Ts* ≤ 1.08, *Ps* ≥ 0.29*)*.

Divergent thinking expectation effects, Experiment 1.

An additional set of exploratory analyses was carried out to assess the effect of participants expectation over their group allocation on divergent outcome measures. We run two sets of *rm*ANOVAs considering that participant's expectation over group-allocation was likely to change for participants over the course study (baseline, acute 1, acute 2). The first analysis was carried out for the baseline and the first acute session (acute 1). The session (baseline, acute 2) was entered as the within-subject factor, the actual experimental condition (i.e. placebo vs psilocybin) and the expected condition at acute 2 (i.e. placebo, not-sure or active)

were entered as the between-subject factors. The same analyses was carried out for the baseline and the second acute session (acute 2). The analyses was carried out with AUT Z-scores in order to more effectively compare changes in AUT as the function of expectation (regardless of AUT item difficulty presented at different sessions). The main effect of expectation was significant for Fluency and Flexibility at Acute 1 session, yet after visual investigation of the **Figure S1 and S2,** the interaction was driven by placebo condition, who reported not to be sure about their group allocation and thus the interaction could not be interpreted as 'expectation effects' and thus is not relevant to current hypothesis. Similarly, the interactions and main effects of expectation were non-significant in the Acute 2 sessions. Finally, as additional control analyses, Pearson correlations were run between the subjective drug strength and baseline corrected divergent scores that reached significance (e.g. originality and originality ratio). We found a non-significant correlation between the subjective microdosing effect and the change in outcome measures, (rs ≤ 0.096, ps ≥ 0.478).

Convergent thinking expectation effects, Experiment 1.

Next, the *rm*ANOVA was re-ran with expectation over own group allocation (e.g. placebo, not-sure or active) added as the between subject factor (2 x 2 x 3 *rm*ANOVA). We found no main effect of session (F (1,49) = 2.649, p = 0.110, $\eta 2p$ = 0.051), or condition (F (1,49) = 0.395, p = 0.533, $\eta 2p$ = 0.008) after controlling for expectation. Crucially, the main effect of expectation (F (2,49) = 0.222, p = 0.802, $\eta 2p$ = 0.009) and the two-way interaction between expectation and session was also not significant, (F (2,49) = 0.013, p = 0.987, $\eta 2p$ < 0.001). The three-way interaction between time, condition and expectation, (F (2,49) = 0.957, p = 0.391, $\eta 2p$ = 0.030) was also insignificant and the key interaction between condition and expectation (F (2,49) = 0.342, p = 0.561, $\eta 2p$ = 0.007), remain insignificant even after controlling for expectation effects. In sum, the results overall suggested that expectation did not significantly moderate the PCT result.

Fig. S1: Acute 1

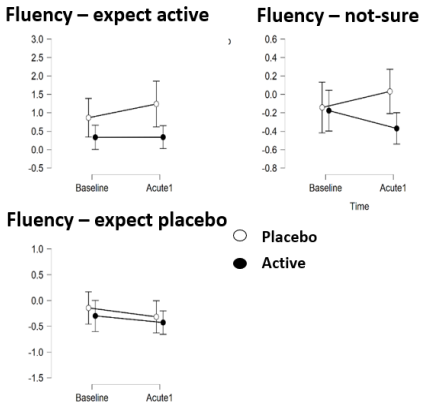


Fig S2: Acute 1

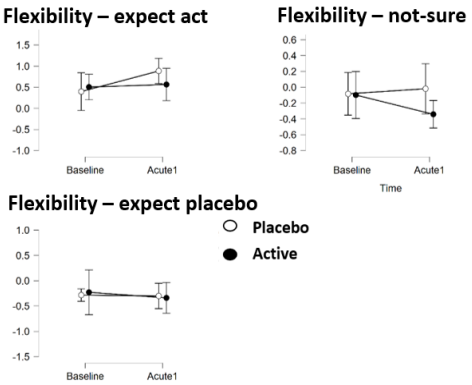


Figure S1. Acute 1 Fluency interaction was driven by participants in a sub-group of participants which were not sure about their group allocation.

Interactions with previous experience

Next, we explored to what degree previous psychedelic experience moderated the drug effects and creativity scores. Interestingly, a significant interaction between condition and previous psychedelic experience was found for the originality ratio, (F (1, 49) = 5.52, p = 0.023, $\eta 2_p$ = 0.101). The drug effects on creativity were mostly driven by participants with prior psychedelic experience for originality and originality ratio scores. Notably, significant baseline difference in elaboration score were found for psychedelic experienced subjects (M=4.5, SD=2.6) as compared to naïve participants (M = 2.6, SD = 2.8); t(51)= 2.38, p=0.021; d=0.66).

The Post-acute effects

The post-acute effects of microdosing on AUT were non-significant as indicated by 2x2 *rm*ANOVAs. The main effects of condition, ($Fs \leq 0.155$, $Ps \geq 0.695$, $\eta 2_p \leq 0.003$) and interaction between time and condition were non-significant, ($Fs \leq 3.236$, $Ps \geq 0.078$, $\eta 2_p \leq 0.056$). Yet, a trend towards significance was observed for the elaboration score (see Supplementary materials, Experiment1, Figure S2).

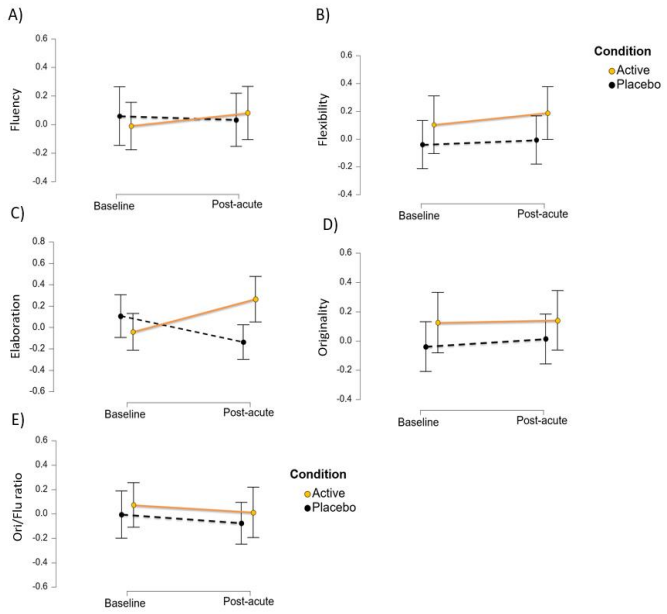


Figure S4. A-E: shows the mean Z-scores for the five divergent scores of the AUT, measured at baseline and post-acutely (~2 days after the last dose) as a function of the group (control – placebo vs. experimental –active psilocybin). Vertical capped lines atop bars indicate the standard error of the mean.

S5. Supplement Experiment 2

Tasks and Questionnaires alterations

Alternate Uses Task (AUT). The AUT measure mirrored the procedure in Experiment 1, requiring participants to think of creative uses of ordinary items. Participants had 5 minutes per item (2 items presented at each session) and as standard, the AUT was rated by two different ratters according to five different variables (for scoring see Experiment 1). Yet, in contrast to previous experiments, the AUT items were presented in different order. For reliability analyses and the list of AUT items used in each experiment, see Supplementary materials (Table S2, Experiment 2).

Picture Concept Task (PCT). Compared to Experiment 1, the PCT measure in Experiment 2 consisted of 15 items presented at baseline and 21 trials presented in the follow-up session in order to increase power of measurement in the experimental phase. Before the analyses, the PCT scores were Z-scored to account for the unequal number of items across the two sessions.

Subjective effects. Participants were again asked to reflect on their subjective microdosing effects during every experimental session. Following procedure in Experiment 1, subjective

effects in terms of microdosing strength and expectations over own group allocation were measured at every follow-up session.

Main Analyses

The Acute 1 session measures were taken after the 6th microdose (previously acute 2) and the Acute 2 session measures were taken after the 10th dose. PCT was tested only twice at baseline and at Acute 1 session, corresponding to testing points in Experiment 1. The analysis strategy followed the procedure at Experiment 1 and data were analyzed by Mixed-design *rm*ANOVAs with session entered as the within-participant factor and group (placebo vs active microdose) as the between-participant factor. The differences in subjective beliefs concerning own group allocation were analyzed with $\chi 2$ tests at every follow-up session. The group differences in the perceived psychoactive microdosing strength were analyzed by *rm*ANOVAs with the session (baseline, Acute 1, Acute 2) entered as the within-participant factor and group (placebo vs active microdose) as the between-participant factor. However, compared to Experiment 1, the exploratory analyses assessing moderating effects of previous psychedelic experience were not carried out because the majority of participants in Experiment 2, had prior psychedelic experience and thus a comparison could not be carried out.

Sign-ups inclusion and exclusion criteria

The screening strategy was identical to Experiment 1. Over 1000 people indicated interest to participate in the workshop on social media yet due to the limited capacity of the workshop, only 100 healthy participants who passed the PSN screening were invited to take part in the microdosing event. From these 100 participants, 96 participants were interested to take part in the associated research. In total 83, healthy participants attended the baseline session at Leiden University. 71 participants attended the Acute 1 session (after 6th dose) and 66 participants attended the Acute 2 session (after 10th dose). Furthermore, another 5 participants had to be excluded as they missed the Acute 1 session or took other psychoactive drugs during the trial. The final AUT sample consisted of 61 healthy participants who completed both follow-up sessions in a good order (30 in placebo and 31 in the experimental group), with a mean age of 27.75 (*SD* = 6.32). In PCT 8 participants had to be excluded, due to incorrect interpretations of the task instructions. The high error rate was likely due to the fact that most participants in the current trial had no prior experience with psychological paradigms. The final PCT sample thus consisted of 55 participants (26 in a placebo and 29 in the experimental group) with an average age of 28.1 (*SD* =6.1).

Notably, due to a software error 8 participants in the AUT task and 6 participants in the PCT task had missing data at baseline while attending the follow-up sessions. Considering

that baseline data served mainly as a control measure to examine possible pre-existing creativity differences among groups, the data for the missing participants were mean imputed (mean per condition) after the control analyses were carried out to preserve experimental power. All participants provided informed consent before the study onset, approved by the local ethical committee (Leiden University, Institute for Psychological Research).

Subjective effects result

Interestingly, even after increasing the active dose's size, participants in the active psilocybin condition did not correctly estimate their group allocation. The chi-square test was not significant at the first, ($\chi 2$ (2, N = 61) = 0.435, p = 0.805, *Cramer's V* = 0.084), nor the second follow-up sessions ($\chi 2$ (2, N = 61) = 0.863, p = 0.650, *Cramer's V* = 0.119). Thus, even after four weeks of dosing participants performed poorly in estimating their group allocation. The result suggests that the blinding manipulation was successful and placebo effects were present even after the prolonged microdosing period see Table 3. (Experiment 2). The group differences in the perceived psychoactive microdosing strength showed a main effect of session, (F (2,118) = 19.510, p < 0.001, $\eta 2p$ = 0.249), indicating that the strength of the dosing subjectively changed across the three measurements. As indicated at Table 4. (Experiment 2) the subjective dose effects again diminished between the workshop (M = 37.13 , SD = 31.88) the first follow-up session (M = 20.96, SD =26) and the last experimental session (M = 16.85, SD = 20.6) ; (t ≤ 5.107 , p < 0.001), regardless of experimental condition. However, the *rm*ANOVA analyses showed a non-significant main effect for condition, (F (1,59) = 0.034, p = 0.854, $\eta 2p$ < 0.001) and an interaction between group and session, (F (2,118) = 0.133, p = 0.876, $\eta 2p$ =0.002). This result indicates that participants did not break blind regarding their experimental condition, since participants perceived comparable psychedelic strength across the active and placebo condition.

Divergent thinking expectation effects, Experiment 2.

An additional set of exploratory analyses was carried out to assess the effect of participants' expectation over their group allocation on divergent outcome measures. We run two sets of *rm*ANOVAs considering that participants' expectation over group allocation was likely to change for participants over the course study (baseline, acute 1, acute 2). The first analysis was carried out for the baseline and the first acute session (acute 1). The session (baseline, acute 2) was entered as the within-subject factor, the actual experimental condition (i.e. placebo vs psilocybin) and the expected condition at acute 2 (i.e. placebo, not-sure or active) were entered as the between-subject factors. The same analysis was carried out for the baseline and the second acute session (acute 2). As shown below significant two-way interaction was found for the session and condition for fluency and flexibility at the Acute 2

session. However, this interaction was driven by a significant difference between the “not-sure” sub-group(See Figure S1 and S2) and thus the interaction could not be interpreted as a result of positive vs. negative ‘expectation effects’.

Fig. S5.1: Acute 2 Fluency

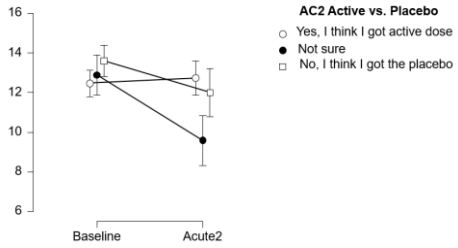
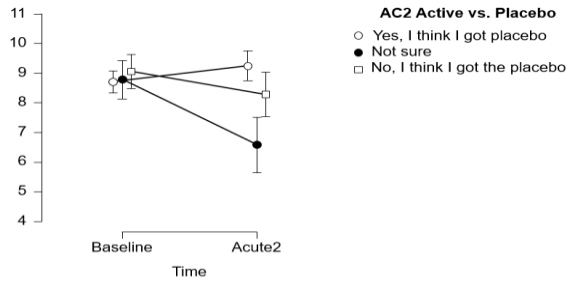


Fig S5.2: Acute 2 Flexibility



Convergent thinking expectation effects, Experiment 2.

Next, we explored the extent to which the participants' subjective expectations may have affected the PCT result. The *rm*ANOVA was re-ran with expectation over own group allocation (e.g. placebo, not-sure or active) at the first follow-up session added as the between subject factor (2 x 2 x 3 *rm*ANOVA). The *rm*ANOVA showed no main effect of session on the convergent thinking (F (1,49) = 0.033, p = 0.857, $\eta 2p$ = 0.001), when controlling for expectation and condition. The main effect of expectation, (F (2,49) = 1.127, p = 0.332, $\eta 2p$ = 0.044),on convergent thinking was not significant. The two-way interaction between session and expectation (F (2,49) = 1.059, p = 0.358, $\eta 2p$ = 0.041) and three-way interaction between expectation, condition and session (F (2,49) = 2.057, p = 0.139, $\eta 2p$ = 0.077), were also non-significant. The main effect of condition, (F (1,49) = 0.265, p = 0.609, $\eta 2p$ = 0.005), and the two-way interaction between condition and session (F (1,49) = 0.016, p = 0.900, $\eta 2p$ < 0.001), remain non-significant even after controlling for expectation effects. The analysis suggests that expectation over own group allocation did not significantly interact with the convergent score.

S6. Supplement Experiment 3

Picture concept task – divergent (PCT-d). Instead of AUT task, we used a divergent version of PCT task, here referred to as the PCT-d. Similarly to previous research (Kuypers et al. 2016; Mason et al. 2019) the PCT-d accounts for both convergent and divergent scores simultaneously. In this task each trial was presented for 2 minutes, instead of 30 seconds as in the previous studies. First participants had a time limit of 30 seconds to find a single

convergent solution by connecting one picture in all the picture rows presented. After completing the convergent task, participants were asked to report as many alternative solutions (e.g., the divergent responses) within the remaining time. Following previous research and our preregistration the dependent variables consisted of one convergent and three divergent parameters (see Fig. 4, Experiment 3). The scoring of the convergent and divergent thinking followed the same scoring strategy as described in Experiment 1 and 2, and was administered by two independent ratters. The PCT-d was administered twice across two testing blocks (i.e., during active psychedelic block and passive placebo block).

PCT-d Task example

Example trial from the Picture Concept Task. The subject had to identify the common association between 1 item from each row. Left image shows the correct convergent solution ("water" by connecting picture 3, 5, 7) and right image shows example of possible divergent solution for the PCT-d ("maintenance" by connecting picture 1, 5, 8).

Pre-processing PCT-d

After completing the convergent answers, participants were asked to look for as many alternative solutions (e.g. the divergent responses) within the remaining time, yet this time participants were allowed to miss one row in order to enhance the divergent thinking process. Thus, if 4 rows of pictures were presented participants had to connect at least 3 rows to gain points for the divergent answers. Following previous research (Mason et al. 2019; Kuypers et al. 2016) and our pre-registered report, one convergent and three divergent parameters were calculated (Mason et al. 2019; Kuypers et al. 2016). The scoring of the convergent and divergent thinking followed same scoring strategy as described in Experiment 1 and 2 and was administered by two independent ratters. Namely the Convergent score (sum of all correct convergent answers), the Fluency score (sum of all divergent responses), Originality score (summing points for unique answers 1% and 5 %) and the Originality ratio (Originality/fluency) across two ratters were averaged after inter-rater reliability was measured. The PCT-d task was carried out once at every dosing block (i.e. week 3 of active vs. week 3 of placebo

dosing). Considering the cross-over design of Experiment 3, and contra-balancing of conditions across participants, PCT-d blocks' difficulty needed to be equal across the two conditions. Thus, a pilot study (N =20) with IRT analyses (i.e. Item response theory) was carried out before Experiment 3, to design similar difficult test-sets for each block. Next, all the participants received identical sets of PCT-d items within each block, and different items were used across the blocks to control for learning effect. The final PCT-d scores were further standardized (Z-scored) per block, to further control for possible deviation in the PCT-d test difficulty.

Sign-ups and exclusion criteria

Prior to the study participants were asked for contra-indications, including a prior diagnosis or family problems with schizophrenia, psychosis, mania, or borderline. We also excluded participants who indicated to have an addiction, who had serious physical health issues (e.g., diabetes or brain injury) and who lacked proficient English language skills (as the study was conducted in English). All participants provided their written informed consent to participate in the study. We provided participants with the guidelines as specified by the researchers involved in this project, asking them to comply with the proposed microdosing schedule for 2 months, self-administering the dose 1.5 h prior to the lab session, to remain blind to their condition, to self-administer at least five of the seven microdoses per block, and to refrain from using other psychoactive substances and medications during the study. The duration of 1.5 h after ingesting the microdosing capsule is based on the observation that the subjective effects of a higher dose of psilocybin and the plasma levels tend to peak around 90 min after intake (Passie et al. 2002). Only participants who complied with behavioral guidelines and completed both testing blocks in good order were included in the final analyses (e.g., active microdosing block and placebo block). Initially, 100 healthy participants who passed screening were invited to join the microdosing workshop. From the 100 participants, only 75 started out with our study and filled in the initial research screening information. Twenty-five participants dropped out after the first block and additional 16 participants dropped out during the second block of testing (i.e., approximately eight weeks after the first dose). In total 34 participants completed both required sessions, but 7 participants were further excluded after screening (5 took other psychoactive drugs and 2 participants self-administered microdoses longer than 2.5 hours before testing). This yield a final sample of 27 healthy participants (13 females) between the age of 20 to 48 with mean age 31.3 years (*SD* = 10.00) from which 14 participants started with the active doses.

Subjective effects- breaking blind

The differences in subjective beliefs concerning own group allocation were analyzed with χ2 tests at each of the two experimental blocks. The chi-square analysis indicated that the placebo manipulation was successful, during the first block of testing (after 7th dose), (X^2 (2, N = 21) = 5.588, p = 0.061, *Cramer's V* = 0.516), however at the following block (14th dose), (X^2 (2, N = 20) = 11.209, p = 0.004, *Cramer's V* = 0.749) participants were breaking blind. Next, the doses' psychoactive strength across the two conditions were analyzed. The independent samples t-tests indicated non-significant difference between conditions at the first block (7th dose) t(32) = 1.40, p = 0.173, *Cohen's d* = 0.541), yet in the second block (14th dose) participants in active condition rated the microdosing effects significantly higher (M = 33.23, SD = 30.09) than in the placebo condition (M = 4.28, SD = 8.01); t (32) = 3.23, p = 0.002, *Cohen's d* = 1.33). Furthermore, Table 5. (Experiment 3) indicates that subjective microdosing effects diminished over time regardless of condition between the workshop and first follow-up session (t (34) = 2.74, p = 0.011, *Cohen's d* = 0.52), which is a finding consistent with the previous two experiments.

Exploratory analyses of subjective effects.

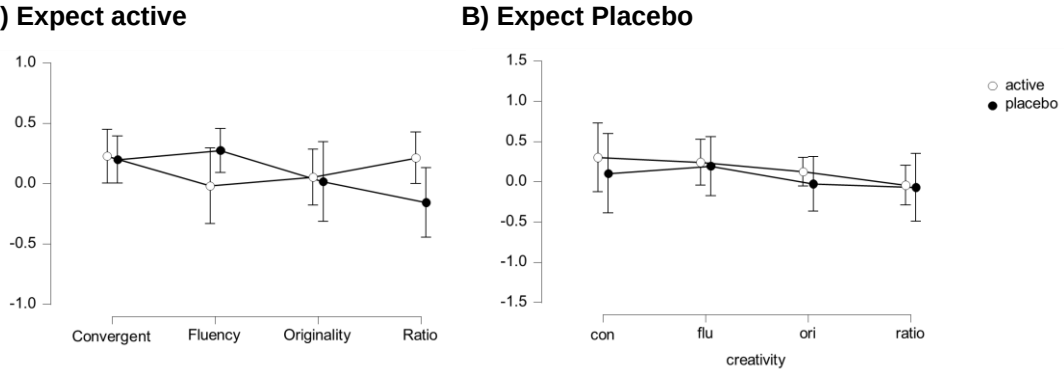
In addition to the pre-registered analyses, exploratory analyses (following the strategy of Experiment 1 and 2) were run to assess the effect of participants' expectations over own group allocation (e.g., active, placebo, not-sure) on the change in the outcome measures. The data were analyzed by 2 x 4 x 3 *rm*ANOVAs with the type of treatment (placebo vs active microdose) and creativity scores (i.e. convergent score, fluency, originality, and originality ratio) as the within-participant factors and expectation regarding group allocation in block 1 and 2 (e.g., active, placebo, not sure) as the between-participant factor. In addition, correlations were run between the ratings of subjective drug strength at each block and the change in the outcome measures.

The 2 x 4 x 3 *rm*ANOVA indicated lack of main effects for expectation, (F (2,17) = 0.274, p = 0.764, *η2p* = 0.031). The two-way interaction between expectation and condition, (F (2,17) = 1.149, p = 0.340, *η2p* = 0.119), and the three-way expectation between expectation, condition and creativity was significant, F (2,17) = 0.834, p = 0.549, *η2p* = 0.089), were insignificant. This suggests that participants' expectations did not significantly interact with the results. However, while controlling for the variance accounted by expectation the key interaction between condition and creativity was rendered insignificant, (F (2,17) = 0.999, p = 0.369, *η2p* = 0.055). The lack of interaction is not surprising considering the fact that condition

and expectation variance significantly overlapped in the second block as the participants were breaking blind.

In order to further inform our results in terms of subjective drug effects and their influence on significant creativity change, Pearson's correlations were run between the change in the originality ratio score (e.i. placebo – active) and the subjective rating of the microdosing strength, at each of the two testing blocks separately. The correlations were not significant (Rs ≤ 0.212, Ps ≥ 0.228) suggesting that the subjective drug effects did not significantly influence the result.

Figure S1. The line graphs represent convergent and divergent creativity scores (mean +/- standard error) for (a) participants expecting to be in active condition (b) participants expecting to be in the placebo condition at Block 2 (when participants were breaking blind).



S7. Mega-analyses

Subjective effects and dose size

The Pearson's correlation between participants weight and subjective microdosing effects was negatively correlated (*r(153) = 0.181, p = 0.025) and* the Pearson's correlation between the relative dose size (weight/dose) and the ratings of subjective microdosing effects was positively correlated (*r(153) = 0.22, p = 0.045)* suggesting mediating effects of weight on the psychoactive microdosing effects.

Step-wise regression data-screening step-wise regression

The data was screened for assumptions of linearity, normality, homoscedastic, and multi-collinearity. In order to run the multiple regression necessary assumptions: independence of

errors, homoscedasticity of residuals (equal error variance), errors are normally distributed, multicollinearity (not highly correlated predictors)

- First the Durbin-Watson case-wise diagnostics was run to detect independence of error residuals (standard residual > 3). The statistics were close to 2 and alpha was not significant suggesting non-significance, indicating that participant data were independent.
- Next, we investigate whether our variables are linearly related to the dependent variable by investigating of residual plots, Q-Q plot and Residual Histograms. Residual histograms and Q-Q plots were appropriate.
- Next we tested that the homoscedasticity of the residuals is equal for all values of the predicted dependent variable. The residual spread indicated equal distribution over the predicted values of the dependent variable suggesting the data are homoscedastic.
- Multicollinearity occurs when one or more independent variables are highly correlated with each other. This leads to the problem of understanding which variable contributes to the variance explained. If the VIF for a factor is near or above 5 – we need to remove redundant factors.
- Thus, first we inspected the strength of the correlation between the continuous IVs as well as examined Tolerances/VIF values. None of the correlation was higher than 0.276 indicating moderate correlations- suitable for regression (considering the recommended threshold of < 0.7).

S6.4. Mega-analyses Hierarchical regression – condition in step 2

Finally, to determine to what degree subjective and demographic factors interact with the creativity scores, we performed series of two-stage hierarchical regressions for each dependent measure to evaluate the nuisance variables' contribution to the creativity scores. The data were entered in a linear hierarchical regression with each of the four creativity scores entered separately as the modeled variable. In the first step, nuisance variables were entered (e.g. trial number, age, gender, weight, subjective drug-strength) in order to control for their variance (*Creativity index= b_{0+} (b_1 × age) + (b_2 × gender) + (b_3 × weight) + (b_4 × experimental trial number) + (b_5 × drug strength) + (b_6 × expectation).* In the second step the condition (active vs. placebo) was included as a regressor (*Creativity index= b_{0+} (b_1 × age) + (b_2 × gender) + (b_3 × weight) + (b_4 × experimental trial number) + (b_5 × drug strength) + (b_6 × group expectation) + (b_5 × condition).*

Appendix E

Supplementary Material for Chapter **6**

S1. Task selection and the measured constructs of each task in Experiment 1

AX-CPT task

The AX-CPT (Servan-Schreiber et al., 1996) is a context processing task that differentiates between proactive and reactive control. The details of this task were based on Hefer & Dreisbach (2017). In the task, participants had to respond to probes on the basis of the preceding cues. In each trial, participants were presented with a cue letter (A or B), followed by a probe letter (X or Y). Participants were required to make a target response only if an AX sequence was presented (an A cue followed by an X probe). All other sequences, meaning an AY sequence (an A cue followed by another probe letter than X), or a BX sequence (a probe letter X following another cue letter than A), or a BY sequence (two different cue- and probe-letters), required a non-target response. In this task, the AX sequence is shown in a higher frequency than the non-target sequences, increasing expectancy following an A cue and increasing the tendency to a prepotent target response when presented with an X probe. AX trials occurred 70% of the time and were randomly alternated with the 3 non-target trial types that each occurred with a frequency of 10%. The distractor letters, B-cues and Y-probes, were randomly chosen from the alphabet, beside the A and X. The cues and probes were presented in the color magenta, with a 28 pt. Arial font, and distractor letters were presented in the color black with a 24 pt. Arial font. Participants were thus told to ignore the black distractor letters as only the colored cue and probe letters were important to the task. All letters were presented in boldface and were centered on a gray background. Participants were told that the cue would be presented for a short amount of time and that they had to react only to the probe. At the beginning of the experiment, participants received written instructions and completed a practice block of 20 trials to familiarize them with the AX-CPT. After the practice trials, participants completed 80 experimental trials. Participants were told to react as quickly and accurately as possible. They were informed about their score at the end of the experiment.

The structure of the trials in the task is shown in figure 1. In the figure, it can be seen that the first cue appeared for 300 ms, followed by a blank screen for 200 ms and three distractor letters, each for 300 ms. Then another blank screen was shown for 100 ms, after which the probe was presented and was left on the screen until the participant gave a response. Participants received a visual feedback message after their response that lasted 1500 ms. The word “correct” was presented after correct responses and the word “wrong” was presented in red after an incorrect response. In between trials, an interval with a blank screen was presented for 500 ms. For the analysis of performance data, practice trials and the first trial of each experimental block were excluded. In addition, error trials were excluded prior to the mean median RT analyses.

Indexes and hypothesis:

The difference between proactive and reactive control is highlighted in the dual mechanisms of control theory (DMC) which differentiates between proactive and reactive control (Braver, 2012; Braver et al., 2017). Usually, there is a preference for one control mode over another, which can be investigated using the AX-CPT. Proactive control is goal directed, leads to preparatory activation, and aims to prevent interference. Participants using proactive control maintain information around the cue, allowing them to prepare a target response. This strategy is beneficial in BX trials but lowers performance (more errors and higher reaction times) on AY trials, because on a BX trial, a non-target response is already prepared, while at AY trials, an incorrect target response is prepared. Participants using reactive control on the other hand, maintain information around the probe, and cue information is less actively maintained. This strategy is beneficial in AY trials, because there will not be a tendency to erroneously give a target response based on the cue information. However, this strategy is costly in BX trials because of the association of the X probe and a target response, which will not be corrected for by the cue-based information. The target response bias that is triggered by the X-probe thus has to be overcome and episodic information about the context has to be retrieved from memory, which is time consuming resulting in higher reaction times.

According to literature, the more cognitive control one exerts, the more persistent they are, increasing proactivity (Hefer & Dreisbach, 2017). Under psychedelics, it can be hypothesized that people exert less cognitive control (Carhart-Harris and Friston, 2021), and according to the Metacontrol model, reduced top-down control is thought to shift the balance from persistence to flexibility (Hommel, 2015), thus it would be expected that people are less proactive. Accordingly, we expected that people use more reactive control under psychedelics and will thus perform better in AY trials, and worse in BX trials. Performance on this task is measured using reaction time and error rates. For the analysis, the practice trials and the first trial of each experimental block was excluded.

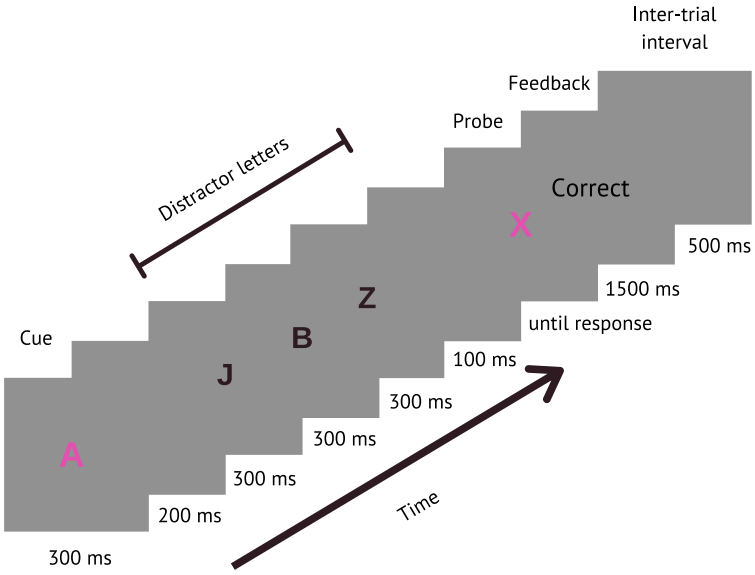
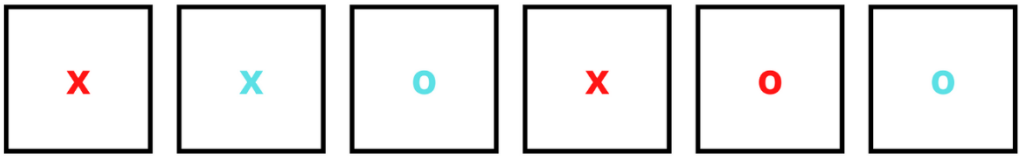


Figure 1. Example of an AX-CPT trial. Reference-back task**.** The reference-back task (Rac-Lubashevsky & Kessler, 2016) is a working memory task which isolates the updating process. There are two different types of trials in this task, namely comparison trials, which involve a matching decision, and reference trials, which involve both a matching decision and updating cost. The details of this task were based on Jongkees (2020). Participants were first presented with a white fixation cross that was centered on a black screen. After 1 second, participants were presented with either an X or an O, in either a blue (comparison trial) or a red color (reference trial), which was presented for 2 seconds, serving as the first reference stimulus. Participants had to indicate whether the stimulus (X or O) they were presented with, was the same as or different from the previous stimulus that was presented in a red (but not blue) color, using the 'q' and 'p' buttons on a QWERTY keyboard. In addition, if the stimulus was both different and presented in a red (but not blue) color, the participant had to update the comparison stimulus to the one being presented.

In all trials except for the first one, the stimulus color (red or blue) and identity (X or O) were randomized, meaning that each trial had a 50% chance of being a reference or comparison trail, a switch or no switch trial (switching refers to switching between the reference and comparison trials) and requiring a 'same' or 'different' response. The stimuli were presented until a response was given. An example of a part of a block is presented in Figure 2. Participants first completed 2 practice blocks, after which they completed 12 experimental blocks of 40 trials each. Participants were told to respond as fast and accurately as possible. Performance on this task was measured by the error rates and reaction times. For the analysis, the first trial of each block was excluded. In addition, for the analysis on the reaction times, error and post-error trials were excluded.

Figure 2. *Reference-Back Task Example with Response, Trial Type and Possible Switching.*



Response	-	Match	Mismatch	Match	Mismatch + update	Match
Trial type	Reference	Comparison	Comparison	Reference	Reference	Comparison
Switching	-	Switch	No switch	Switch	No switch	Switch

This task also distinguishes between different subprocesses that take place in working memory. The subprocesses investigated in this study are: the updating cost (the difference between reference and comparison trials), updating cost without a switch (the difference between reference and comparison trials without a switch), gate opening (the switch cost in reference trials, calculated by taking the difference on reference switch trails and reference no switch trials, as they both require an open gate to update WM, but switch trials also require opening the gate after it was closed on the previous trial), gate closing (switch cost in comparison trails, calculated by taking the difference between comparison switch and comparison no switch trials, as they both require a closed gate for WM maintenance, but switch trails also require closing the gate after it was opened on the previous trial), and switch cost (switch versus no switch).

Indexes and hypothesis:

This task also distinguishes between different subprocesses that take place in working memory. The subprocesses investigated in this study are: the updating cost (the difference between reference and comparison trials), updating cost without a switch (the difference between reference and comparison trials without a switch), gate opening (the switching cost in reference trials, calculated by taking the difference on reference switch trails and reference no switch trials, as they both require an open gate to update WM, but switch trials also require opening the gate after it was closed on the previous trial), gate closing (switch cost in comparison trails, calculated by taking the difference between comparison switch and comparison no switch trials, as they both require a closed gate for WM maintenance, but switch trails also require closing the gate after it was opened on the previous trial), and switch cost (switch versus no switch).

While persistence prevents distractions, flexibility keeps working memory up to date. As such it can be hypothesized that active microdosing conditions would enhance updating (flexibility) associated with gate opening in REF, while maintenance (persistence) associated with gate closing in REF would be hindered. In addition, switching between task sets may also reflect cognitive flexibility, as participants have to respond differently to the same input (Kessler, 2017), therefore, we would expect people under the influence of microdosing to perform better on switch trials than people who are not under the influence of psychedelics.

Remember-Know task. The remember-know task (Tulving & Annis, 1985) is a long-term memory task to study subjective experiences of recognition. The details for this task were based on Keizer et al. (2010). The experimental stimuli were 156 line-drawing pictures, divided in three lists of 52 items, drawn from Snodgrass and Vanderwart (1980), Berman et al. (1989), and Cycowicz et al. (1997). Participants first competed a practice block, after which they carried out two experimental blocks. Each block consisted of 2 parts. The first part was the encoding phase with 36 stimuli and 4 extra filler stimuli (the first and last to stimuli of each encoding block, used to avoid primacy and recency effects), presented in red or green (randomly and equally presented). Participants were told to memorize the stimulus and its color and had to press a left- or right-side button (which was counterbalanced between participants), depending on the color of the stimulus. Each stimulus was presented for 700 ms, followed by a blank interval of 1300 ms until the next stimulus was presented. The second part of the block was the retrieval phase with 26 stimuli presented in black-and-white, of which 14 were new, and 12 were old (6 previously presented in red, and 6 in green). Participants were asked whether the stimulus they were presented with was “old-remembered”, “old-know”, or “new”, by pressing on of three response buttons (which was counterbalanced between participants). The three response options were shown below the drawing. On the one hand, when a “new” response was given, a blink interval was presented for 1000 ms, after which the next drawing was presented. On the other hand, when an “old-remembered” or “old-know” response was given, the drawing would stay on the screen, and participants were asked in what color (red or green) the drawing was shown in the encoding phase. The response options (red or green) were displayed below the drawing and participants had to respond by pressing a left or right button, which was counterbalanced between participants. After participants have given a response, a blank interval was presented for 1000 ms, after which the next drawing was presented.

The encoding phase is used to assess recognition memory, in which “old-remembered” is believed to assess recollection and “old-know” is believed to assess familiarity (Burgess & Ali, 2002). Recollection refers to the conscious recognition of an event, including contextual information, while familiarity refers to a weaker recognition, not including contextual

information. Memory is expected to be worse under the influence of psychedelics. Therefore, we would expect worse performance on the old pictures. Performance is measured by the amount of correct and incorrect categorizations. The filter stimuli form the encoding blocks were excluded from the analysis.

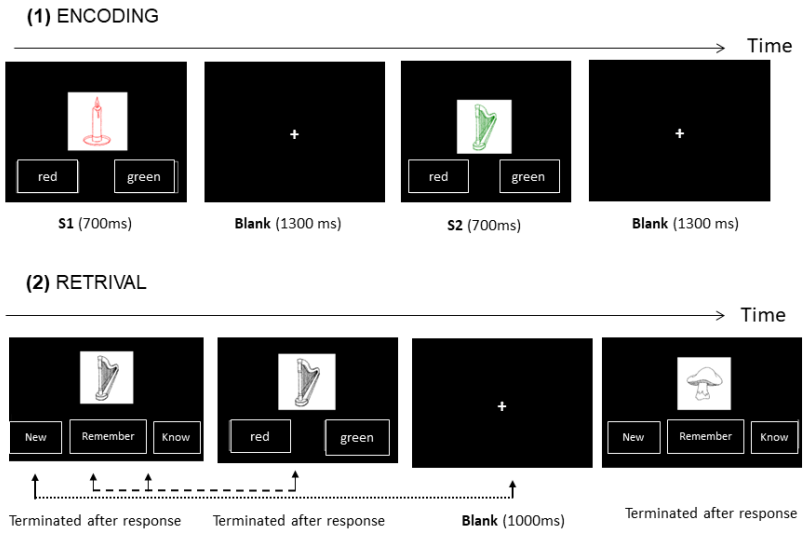


Figure 3. Long-term memory task. Schematic representation of the sequence of events occurring on each trial of the two consecutive phases (encoding and retrieval) of the long-term memory task.

Multi-Armed bandit task. The multi-armed bandit task is a task that can be used to measure persistence by exploring the tradeoff between exploration versus exploitation (Cohen et al., 2007; Gittins et al., 2011). The details of this task were based on Mekern et al. (2019). In the task, participants had to choose whether they wanted to exploit their current alternative or explore for possible better alternative, that might however, also be worse. In sum, the participant had to allocate a limited set of resources between different choices, to maximize the expected gains. This is a typical reinforcement learning problem, where each choice provides a random reward. The participants had to choose between different slot machines and attempt to maximize their rewards by exploiting the machine(s) with the highest expected payoff that is known. Alternatively, they could choose to explore values of the other machines and acquire new knowledge. In this task, participants played four slot machines and their goals was to gain as many points as possible (Daw et al., 2006). There was one slot machine in each corner of the computer screen and participants had to choose an arm by pressing the keys Q, W, S, or A. An example trial can be seen in figure 4. The task consisted of 200 trials, and the payoff of each arm changed gradually during the task. In this way, participants constantly had to adjust their explorations vs. exploitation strategy to maximize their gains (Daw et al., 2006; Jepma et al., 2010; Jepma & Nieuwenhuis,2011). The reward of each trial

was presented for 800 ms, and the total reward was always displayed. The slot machines stayed on the screen until the participant responded. The participants' performance was measured by calculating how many times participants switched or stayed, and whether they won or lost something by staying or lost something by switching. Under the influence of psychedelics, people are thought to behave more flexibly, and therefor are expected to show more exploration behavior on this task. We were therefore mainly interested in whether the number of switches was higher under the influence of psychedelics and see whether this influenced performance on this task.

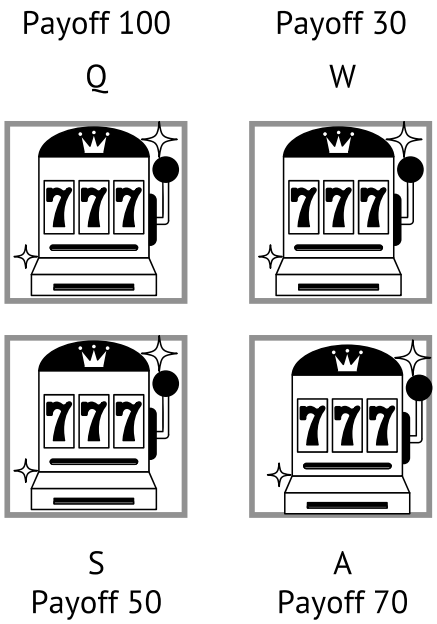


Figure 4. Multi-Armed bandit task**.**

Reading the mind in the eyes task (RMET)

Participants were tested three times on a computerized version of the RMET – revised (Baron-Cohen et al. 2001), using the E-Prime 2.0 software system (Psychology Software Tools, Inc., Pittsburgh, PA). The task consisted of 36 black-and-white photos of a human's eye region, presented one by one along with four adjectives (one target word and three foil words) arranged around the eye region. On each testing day, participants were instructed to choose which of the four words better described what the person in the picture was thinking or feeling, by clicking on it with the mouse. The 36 target pictures were preceded by an additional picture that served as an example. Participants were asked to respond as fast as possible, but no time limit was imposed. The task w in English, and whenever necessary participants were allowed to consult a glossary including a brief definition of each word. Following previous studies (Domes et al. 2006; Guastella et al. 2010; Baron-Cohen et al. 2015; Colzato et al. 2017a), the 36 items of the RMET were divided in two subsets of easy and difficult items. The

two subsets of stimuli were based on the median-split of item difficulty performed in a pilot study carried out by Colzato et al. (2017a). Performance on this test is measured by computing the percentage of correct responses as a function of item difficulty.

S2. Subjective measures Experiment 1

Multidimensional psychological flexibility inventory (MPFI)

Multidimensional psychological flexibility inventory (Rolffs, Rogge, & Wilson, 2018) was used to assess global cognitive flexibility. MPFI is a 60-item questionnaire consisting of 12 subscales based on the Helaflex model (Hayes et al., 2011). The questionnaire measures 6 domains of flexibility and 6 domains of inflexibility. The 6 core domains of flexibility include: acceptance, present moment awareness, self-as-context, defusion, and committed action. The 6 domains of inflexibility include: experiential avoidance, lack of present moment awareness, self as content, fusion, inaction, and lack of contact with values. The items, such as "I opened myself to all my feelings, the good and the bad", are rated on a 6-point Likert scale ranging from 1 (*never* or *never true*) to 6 (*always* or *always true*). The global composite of flexibility can be calculated by averaging the average of the sum of 6 flexibility subscales. Similarly, the global composite for inflexibility was calculated in the same manner (Rolffs, Rogge, & Wilson, 2018).

Warnick Edinburgh Mental Well-being (WEMWBS)

Also known as the Core Questionnaire for Well-Being, the WEMWBS (Tennant et al., 2007) is a 14-item questionnaire that assesses subjective well-being and psychological functioning. The items, such as "I've been dealing with problems well", are rated on a 5-point Likert scale ranging from 1 (*none of the time)* to 5 (*all of the time)*. The WEMWBS is a psychometrically robust scale with no ceiling effects in population samples. Further, it demonstrates good content validity and high test-retest reliability (Stewart-Brown et al., 2011).

Affect grid

Also known as the Affect grid, the mood grid is a single-item self-report scale that assesses how the participant is feeling at a given moment (Russell, Weiss, & Mendelsohn, 1989). The two-dimensional 9x9 grid is plotted with the x-axis indicating the extent of pleasure and the y-axis indicating the extent of arousal. The axes divide the grid in the following quadrants: stress, excitement, boredom, and relaxation. Participants were asked to mark the grid based on how they are currently feeling, and mood is then assessed based on the placement of the mark. The mood grid demonstrates good convergent and discriminant validity, as well as high reliability. More importantly, it takes approximately one minute to complete (Russel et al., 1989).

Subjective experience questionnaire (SEQ)

Due to the lack of previous research on perceived drug effects, a 9-item questionnaire was developed specifically for this experiment. The SMEQ captures alterations in aspects of mood, visual perception, and cognition after microdosing. This questionnaire was design to assess subjective effects which could potentially indirectly influence creativity measures. The questionnaire included items such as: *"Under the influence of a microdose I have been feeling distracted".* Participants made their responses on a sliding Likert-like scale ranging from 1 (*"exceptionally less than normal"*) to 7 (*"exceptionally more than normal"*), where 4 referred to no change from normal.

The nine-item questionnaire administered at Experiment 1 at the two follow-up sessions.

1. Please answer the following questions: - To what extent are you experiencing mood swings?
2. Please answer the following questions: - How anxious do you feel?
3. Please answer the following questions: - Are you thinking clearly?
4. Please answer the following questions: - To what extent are you feeling distractible?
5. Please answer the following questions: - Are you feeling disoriented?
6. Please answer the following questions: - Do you notice a change in your depth perception?
7. Please answer the following questions: - Is your color perception different?
8. Please answer the following questions: - Do you experience a difference in visual clarity?
9. Please answer the following questions: - Do you feel self-confident?

Subjective drug effects

Participants were asked to describe their microdosing experience during acute 1 and acute 2 in order to capture their subjective microdosing experience. More specifically, in acute 1 participants were asked to "describe the main effect(s) you felt during your first microdose, also mention if you did not notice any changes" and "do you think that you are taking an active microdose?" In acute 2, participants were asked to "describe the main effect(s) you have been feeling in regards to microdosing over these past 2 weeks" and "do you think that you are taking an active microdose? If yes, you feel symptoms such as…" They are free to express what they are feeling in the present moment and these responses were recorded as text data.

Text analysis

In baseline and acute testing sessions, participants were asked to report on their subjective microdosing experience. Verbal data were analysed by classifying whether reported drug

effects were positive or negative; and cognitive, emotional, bodily, and/or social. Further, sub-categories within cognitive and emotional effects were created. Within *cognitive*, self-reflection, persistence, and cognitive flexibility were included; and within *emotional*, mood and theory of mind were included. All drug effect classification variables were coded as binary outcomes. While acknowledging that effects are not mutually exclusive, data was coded in an additive manner. For instance, a participant could experience increased efficiency in finding solutions, agitated mood, and increased visual acuity. “Increased efficiency in finding solutions” is coded as positive, cognitive, and flexibility, while “agitated mood” is coded as negative, emotional, mood, and “increased visual acuity” is coded as positive and bodily. Each quote is analysed in segments in order to avoid only accounting for the most salient drug effects.

S3. Task selection and the measured constructs of each task in Experiment 1

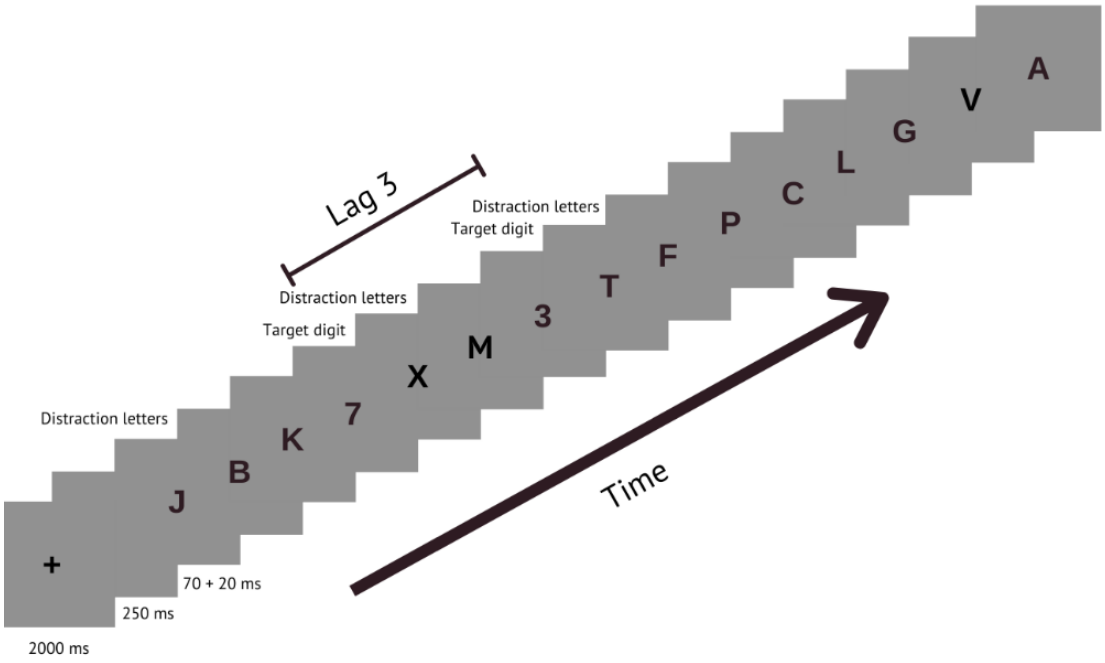
Attentional-blink task. The attentional blink means that there is a decreased performance for reporting a second target when it is presented shortly after a first target (Raymond et al., 1992). The attentional-blink task can be used to assess cognitive flexibility and measures the efficiency with which attention is allocated. Participants had to identify two digits that were presented in a stream of letters (distractors), appearing rapidly after one another. The first target (T1) is often accurately reported, but the second target (T2) is more difficult to report if presented shortly after the first target (within 200 – 500 ms). With better allocation of attentional resources, participants show more parallel processing of the targets, which leads to a better performance in the attentional blink task.

The specifics of this task were based on Trutti et al. (2019). At the beginning of each trial, a fixation cross was presented for 2000 ms, followed by a blank interval for 250 ms. Next, a rapid serial visual presentation (RSVP) was presented, consisting of 15 letter/digit items. Each stream consisted of 2 target digits, ranging from 1-9, and 13 distractor letters, in random fashion and without replacement, ranging from A-Z, excluding I, O, S, and Z, as these letters too closely resemble some of the digits. Each item was presented for 70 ms with an inter-stimulus interval of 20 ms. All items were presented in black in a 16-point Times New Roman font, centered on a gray background.

The task consisted of 3 conditions, namely a lag 1, a lag 3, and a lag 8. An example of a lag 3 trial can be seen in figure 1. The position of T1 (target 1) was varied in the RSVP stream to being the 3rd, 4th or 5th item, to reduce predictability. T2 was presented directly after T1 (lag 1), after two distractors (lag 3), or after seven distractors (lag 8). Participants had to report both targets directly after the RSVP, by pressing the correct number keys. The order of the numbers was not important.

Participants first completed a minimum of 18 practice trials, after which they completed two experimental blocks of 72 trials each (each of the 3 lags 24 times). Performance was measured with the accuracy on T2. Under psychedelics, we expected participants to perform better in the lag 3 condition. Usually, performance is impaired at 270 ms, but as people under the influence of psychedelics were expected to behave more flexibly, we expected a better allocation of attentional resources as they could switch the allocation more quickly, leading to a better performance on this condition. On the other conditions, we did not expect any differences between the experimental and control group.

Figure 1.



N-back task *Example trial of the attentional blink task with a 300 ms lag.* The n-back task is often used to investigate working memory (Jonides et al., 1997; Kirchner, 1958; Pelegrina et al., 2015). The details of this task were based on Colzato et al. (2013). In the task, participant had to decide for each trial whether the item they are presented is the same as or different from the one presented in *n* trials before. The *n* is typically varied between one and three, but in this case 0, 1, and 3 were used. This is in order to manipulate the storage and complexity of updating (processing load), from no load to a high working memory load. The 0-back condition was used as a 'baseline', without working memory load, the 1-back condition was used to assess the low working memory load (can rely on immediate perceptual priming) and lastly the 3-back condition was used to assess a high working memory load (requires online monitoring and updating of WM, which is cognitively demanding (Kane et al., 2007).

The items that are presented were letters and participants had to react to the letter when it was the same as the *n*-th letter before the stimulus letter (in the 0-back, this was a pre-specified letter). Participants thus had to constantly remember the last *n* series of letter, which rapidly changed meaning that the participants had to maintain and manipulate information in the working memory (Colzato et al.,

2013). The difficulty increases with an increasing *n*, because longer sequences have to be remembered. Participants had to press a 'Z' or an 'M' key on the QWERTY keyboard to indicate their responses, which was counterbalanced between subjects. The stream of letters (consisting of either of these letters: B, C, D, G, P, T, F, N, L) was presented for 1000 ms. 33% of the trials consisted of targets, the rest were non-targets. Stimuli presentation was pseudo-randomized to prevent lure trials, where non-targets match a recent letter in the stream of letters, but not the target, eliciting more false alarms and misses than non-lure non-targets (Kane et al., 2007). An example of a 3-back trial of the N-back task can be seen in figure 2.

In the task, participants received instruction after which they had to do some practice trials. They then had to acknowledge that they understood the task. After the practice trials, the experimental trials were given, which consisted of 102 0-back trials, 104 1-back trials, and 108 3-back trials. In each *n*-back condition, there were 42 targets, and the remaining trials consisted of non-targets. The presentation of stimuli was pseudo-randomized to avoid lure trials (non-target letters that match a recent, but not current letter of the N-back), as these elicit more false alarms and misses than non-lure trials (Kane et al., 2007). Performance was measured with accuracy, meaning the amount of 'hits' (when participants pressed the key when a target letter was presented) 'misses' (when participants did not press the key when a target letter was presented), 'false alarms' (when participants pressed the key when no target stimulus was presented) and 'correct rejections' (when participants did not press the key when no target stimulus was presented'.

In addition, d'prime was used as a target sensitivity index as suggested by Jongkees et al. (2017), to accurately assess working memory. The index enables discrimination between targets and non-targets. The measure combines hit and false alarm rates, providing an index of the ability to discriminate targets from non targets. D' prime is calculated by subtracting the standardized false alarm scores from the standardized hit scores. Higher d' prime scores signal selective, correct reporting of the targets and thus better WM performance. Each participant had nine d' prime values from the three conditions (0-back, 1-back and -back) and the three lab sessions (baseline, acute 1, and acute 2). Under the influence of psychedelics, working memory performance is expected to be worse. Therefore, people that are under the influence of psychedelics (experimental group) are expected to perform worse on this task and make more errors, and thus have a lower d' prime than people who are not under the influence of psychedelics (placebo group), especially under a higher working memory load. In addition, we expect that people are under the influence of psychedelics are more impulsive and will therefore have more false alarms than people who are not under the influence of psychedelics. Overall, we expect performance to be worse, meaning more misses and false alarms, and less hits and correct rejections than people who are not under the influence of psychedelics.

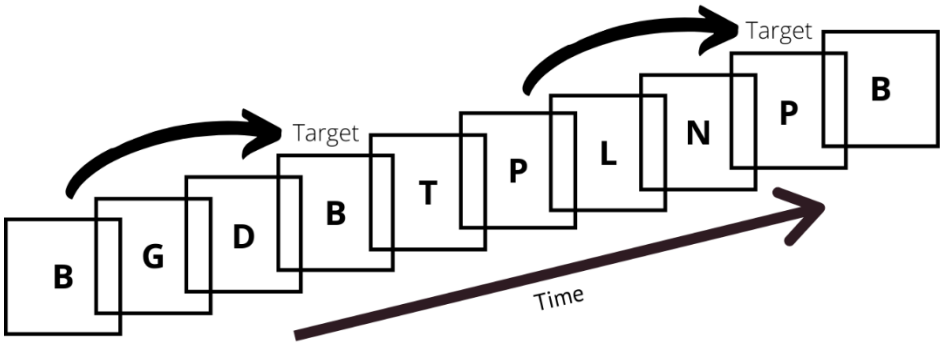


Figure 2. *Example trial of a 3-back trial.*

Trust game. The trust game can be used as a measure of personal trust of another person. For the trust game (Berg et al., 1995), participants had to come to the lab in dyads that were unacquainted and from the same sex (Sellaro et al., 2014; Sellaro et al., 2015). After they were introduced to each other by the experimenter, they were seated in separate cubicles. The participants were led to believe that in the game, one of them would play the role of a trustor and the other would play the role of the trustee, while in reality, all participants played the role of a trustor. They were given €5,-, which they could either keep, partially transfer, or fully transfer to the trustee, whom they believed to be the other member of their dyad. The participants were told that the money they transferred to the trustee would be tripled, after which the trustee had to decide if and how to share the new amount of money. Thus, the trustor could gain more money by transferring money to the trustee, but only if the trustee would give enough money back. The amount of the money that is transferred by the trustor to the trustee is an indicator of personal trust (Camerer, 2003). The trust game can be seen in figure 3. We expected increased trust under the influence of microdosing, as cognitive flexibility was previously linked to prosocial behavior.

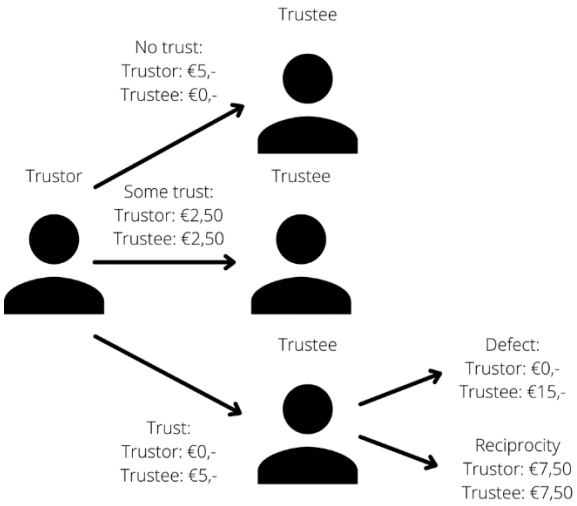


Figure 3. *Trust game examples of with no trust, some trust and full trust from the trustor and the possibility for defect or reciprocity from the trustee.*

S4. Subjective measures Experiment 2

IOS. Participants had to fill in the IOS (Inclusion of Other in the Self scale) (Aron et al. 1992). In the IOS scale, participants had to select the picture that best described their relationship with the other member of their dyad, using a set of Venn-like diagrams as can be seen in figure 4. Each diagram represents different degrees of overlap of two circles. The degree of overlap increases linearly in seven steps, indicating increased closeness to the other with each step. We expected to observe increased self-reported closeness to others under the influence of microdosing, considering that previous research shows increased self-other overlap with increased flexibility (Colozato et al., 2013).

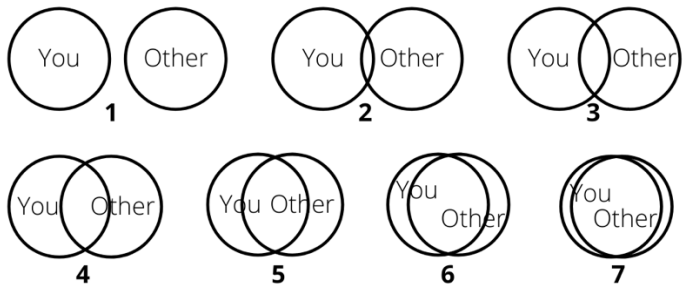


Figure 4. *IOS scale with seven Venn-diagrams representing closeness to the other person, with an increasing overlap in seven steps.*

Cognitive flexibility Inventory (CFI)

The CFI (Dennis & Vander Wal, 2010) is a 20-item questionnaire that assesses the ability to successfully challenge and update maladaptive thoughts with adaptive thinking. It assesses 3 domains: the tendency to perceive difficult situations as controllable; the ability to perceive multiple explanations for life occurrences; and the ability to generate multiple alternative solutions to difficult situations. Participants are asked to report the extent to which they agree or disagree with the items on a 7-point Likert scale, ranging from 1 (strongly disagree) to 7 (strongly agree).

Freidburg Mindfulness Inventory (FMI)

The FMI is a 14-item questionnaire that measures general mindfulness such as being present in everyday moments (Walach, Buchheld, Buttenmuller, Kleinknecht, & Schmidt, 2006). The purpose of this inventory is to characterise the subjective experience of mindfulness; thus, participants are reminded that there is no objective correct or incorrect answers and are urged to answer as honestly as possible. Items, such as "when I notice an absence of mind, I gently return to the experience of the here and now", are rated on a 4-point Likert scale, ranging from 1 (rarely) to 4 (almost always). However, the item, "I am impatient with myself with others", is inversely scored. For the inversely scored item, 1 indicates *almost always* and 4 indicates *rarely*.

Self-Compassion Scale (SCS)

The SCS (Neff, 2016) is a 26-item questionnaire that measures six domains of self-compassion. These six domains include: self-kindness (e.g. "I try to be loving towards myself when I'm feeling emotional pain"); self-judgment (e.g. " I am disapproving and judgmental towards my flaws"); common humanity (e.g. "my failings are part of the human condition"); isolation ("when I fail I feel more separate from the rest of the world"), mindfulness (e.g. "I try to take a balanced view of the situation"), and over-identification (e.g. "when I'm feeling down I tend to obsess over everything that's wrong"). All items are rated on a 5-point Likert scale, ranging from "never" to "almost always". Mean scores are calculated for each domain and the grand mean serves as an indication of overall self-compassion.

Positive and Negative Affect Scale (PANAS)

The PANAS (Watson, Clark, & Tellegen, 1988) is a widely used measure designed to capture state affect, trait affect, and emotional fluctuations through time. The questionnaire consists of 20 items, in which 10 items assess the domain of positive affect (e.g., interested, determined) and 10 items assess the domain of negative affect (e.g., hostile, ashamed). All items are rated on a 5-point Likert scale, ranging from 1 (very slightly or not at all) to 5 (extremely). The sum of each domain ranges from 10 to 50, where higher scores of positive affect indicate higher levels of positive affect and higher scores of negative affect indicate higher levels of negative affect.

Appendix F

Supplementary Material for Chapter 8

S1. Principle component analyses results.

In order to establish the individual differences in the impulsive-compulsive spectrum, the Principal-component analysis (PCA) was first used to reduce the S-UPPS questionnaire data with five sub-scales down to leading principal components (Promax with Kaiser Normalization). A two-components (Figure 3) have been originally retained due to having an eigenvalue greater than 1 (eigenvalue = 1.10), which explained 60.6% of the variance in the data. These two components carried eigenvalues of 1.85 and 1.18, respectively. Finally, the communalities were all above .3, further confirming that each item shared some common variance with other items.

Two possible interpretations of the Pattern Metrix (Table 1) were considered. First, we considered the possibility that one component represents the underlying constructs of impulsivity and other compulsivity. Yet, as shown in Table 1, the subscales loading highly on the first component were Sensation Seeking, Positive Urgency and the Lack of Premeditation. These subs scales include questions such as “I like taking risks” and “I tend to act without thinking when I am really excited” as well as opposite scores for questions such as “I like to stop and think things over before I do them”. Impulsivity is characterised by risk-taking and insufficient information sampling before taking an action (Clark et al, 2006), which is reflected in the first component very well. Compulsivity is characteristic of repetitive behaviours that often lack an adaptive function, which, however, did not reflect the second component well.

The subscales loading highly on the second component were Negative Urgency and Lack of Perseverance, which included questions such as “When I am rejected, I often say things that I later reread” and reverse scores for questions such as “once I get going on something I hate to stop”. Notably, the subscale Lack of Premeditation cross-loaded almost equally high on both components. Perfectionism and high control of thoughts are typical markers of compulsivity (in OCD, Autism REF) and thus second component does not seem to capture compulsivity phenotype. Instead component 2 can be more accurately interpreted as specific sub-type of impulsivity less related to pre-potent motor dis-inhibition as compare to component 1.

Table 1.

UPPS Pattern Matrix		
	Component	
	Component 1	Component 2
UPPS Negative Urgency	0,092	**0,452**
UPPS No Perseverance	-0,167	**0,889**
UPPS No Premeditation	**0,488**	0,441
UPPS Sensation Seeking	**0,862**	-0,178
UPPS Positive Urgency	**0,882**	0,089

Note: UPPS - Positive Urgency Impulsive Behaviors Scale, Extraction Method: Principal Component Analysis. Rotation Method: Promax with Kaiser Normalization.[a]

The second alternative interpretation was that impulsivity and compulsivity are two opposing poles of a single dimension (Robbins et al., (2012; 2018). As such one of the components represents a spectrum - high scores reflect impulsivity and lower scores compulsivity. Indeed, previous research which utilized validated impulsivity (S-UPPS) and compulsivity measures (OBQ-44, IUS) within a large sample of clinical and sub-clinical OCD populations showed a very similar pattern matrix to the current study. In the study by Prochazkova et al. (2018) two-factor loading was derived representing the impulsivity compulsivity component. Relevantly, the high impulsivity component was the best determined by high loadings of Positive Urgency, Sensation Seeking and lack of Premeditation similarly to the component 1 in the current Pattern Metrix. Importantly, the impulsivity component in the study by Prochazkova et al. (2018) was negatively cross-loaded with validated compulsive sub-scales. More specifically ratings of perfectionism, high control of thought and responsibility measured by validated OBQ-44 negatively cross-loaded with Positive Urgency, Sensation Seeking and lack of Premeditation.

Arguably in the current Pattern Metrix, either of the two components may be viewed as an impulsivity compulsivity scale, assuming that impulsivity and compulsivity are indeed anticorrelated constructs (Chmaberlain et al., 2018). Yet, considering that high loadings on component 1 (Positive Urgency, Sensation Seeking and lack of Premeditation) were previously shown to be indicative of low compulsivity scores, the first factor was selected as an index of impulsivity-compulsivity spectrum in the current study. Moreover, the first component showed anti-correlation with the construct of perseverance which is generally a mark of a compulsive trait and thus fits well with the theoretical framework of single impulsivity-compulsivity spectrum. In sum, the single factor solution was selected as consistent with

previous literature as well the theoretical framework of the study. PCA also served to reduce multicollinearity, and subsequent overrepresentation during regression analyses.

S2. Regression analyses

Assumptions test and Model Fit

The assumptions of linearity, normality, homoscedasticity were met and no multicollinearity was present. No outliers were detected. As mentioned above only the originality score of the AUT was used, in order to avoid multicollinearity with other subscales.

The ANOVA table (Appendix B) revealed that the model significantly predicted individual scores on the IMP/COMP scale, $F(4,28) = 5.948$, $p = .001$, adj. $R^2 = .382$. Table 1. shows the regression coefficients, standard errors and t statistics at an alpha of .05. As shown in Table 1, Stroop Error Rate was found to be unrelated to IMP/COMP, $t(39) = -1.094$, $p = .283$.

Table 1. Regression coefficients, standard errors and t statistics

Coefficients							Collinearity Statistics	
Model		Unstandardized	Standard Error	Standardized	*t*	*p*	Tolerance	VIF
1	(Intercept)	0.076	0.142		0.530	0.600		
	Stroop RT	0.607	0.168	0.564	3.611	0.001	0.791	1.265
	Stroop ER	-0.165	0.151	-0.168	-1.094	0.283	0.817	1.224
	Appr/Avoid	-0.443	0.188	-0.351	-2.305	0.029	0.834	1.199
	Originality	0.371	0.150	0.365	2.480	0.019	0.892	1.121

Table 2. Part And Partial Correlations

Model		Partial	Part
H_1	RT Stroop	0.564	0.502
	Error rate Stoop	-0.203	-0.152
	ApprOverAvoid	-0.399	-0.320
	Originality	0.424	0.345