MILESTONE Data Analysis Plan

Study Information

1. **Title:** fMRI investigation of the neural mechanisms of Emotional Cognitive Bias Modification as an adjunct therapy to SSRIs in depression (MILESTONE)

   Short title: Emotional Cognitive Bias Modification in Depression (MILESTONE RCT)

2. **Authors:** Dr Charlotte Crisp, Mrs Alison Burns, Miss Rumeysa Kuruoglu, Dr Sean James Fallon, Dr Jennifer Ferrar, Prof Nicola Wiles, Prof David Kessler, Prof Ian Penton-Voak

3. Description (optional)

Depression constitutes a substantial burden of disease on society, with the DoH estimating economic costs £7.5 billion annually in the UK. Antidepressant drugs are effective, but up to 50% fail to respond to treatment and relapse rates are high. We have developed a Cognitive Bias Modification (CBM) therapy that modifies emotion perception, and targets similar neural processes to SSRIs. Pharmacological antidepressant treatment is most effective when combined with other therapies, but there are few mechanistic studies of adjunctive treatments. We will investigate whether paired psychological and antidepressant treatment have combinatory effects on neural mechanisms that are associated with, and predict recovery from, low mood. Here, using functional Magnetic Resonance Imaging (fMRI), we evaluated this question in patients with depression that had started taking SSRIs in the last 6 months. Participants were randomised to receive an 'active' or 'sham' version of CBM therapy involving training to shift emotion perception of facial expressions in five online sessions over one week. Two weeks after randomisation, participants underwent fMRI scanning during an emotional processing task (viewing emotionally valenced faces). Participants were also followed up at six weeks to assess mood. The trial protocol paper is registered on ISRCTN with the trial registration number 37448835: https://doi.org/10.1186/ISRCTN37448835 .

4. Hypotheses (required)

4.1. List specific, concise, and testable hypotheses. Please state if the hypotheses are directional or non-directional. If directional, state the direction. A predicted effect is also appropriate here. If a specific interaction or moderation is important to your research, you can list that as a separate hypothesis.

4.2. Example: If taste affects preference, then mean preference indices will be higher with higher concentrations of sugar.

The study set out to investigate whether a remotely delivered Cognitive Bias Modification (CBM) psychological therapy, has the potential to be an adjunct therapy to SSRI medication in depression. The aim is to conduct a Randomised Controlled Trial to establish whether there are any differences in the neural correlates of emotional processing following treatment with SSRI between active CBM therapy and sham CBM therapy in patients with depression, recruited through primary care or from the general population. The hypotheses are:

1) People in the active CBM therapy condition will positively change their behavioural responses to ambiguous emotional facial expressions (post-training balance point and learning rate) compared to the sham CBM therapy condition, and this will be evident at 2-week and 6-week follow up.
2) fMRI activation in bilateral amygdala in response to happy vs. sad faces will be greater in patients receiving the active vs. sham CBM therapy.
3) fMRI activation in bilateral amygdala in response to happy vs. fear faces will be greater in patients receiving the active vs. sham CBM therapy.
4) fMRI activation in bilateral amygdala plus medial Prefrontal Cortex (mPFC) in response to happy vs. sad and happy vs. fear will be greater in patients receiving the active vs. sham CBM therapy.

Design Plan

In this section, you will be asked to describe the overall design of your study. Remember that this research plan is designed to register a single study, so if you have multiple experimental designs, please complete a separate preregistration.

5. Study type:

- Experiment - A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.

6. Blinding describes who is aware of the experimental manipulations within a study. Mark all that apply:

- For studies that involve human subjects, they will not know the treatment group to which they have been assigned.

- Personnel who interact directly with the study subjects (either human or non-human subjects) will not be aware of the assigned treatments. (Commonly known as "double blind")
- Personnel who analyze the data collected from the study are not aware of the treatment applied to any given group.

7. Is there any additional blinding in this study?

No.

8. Study design:

8.1. Describe your study design. Examples include two-group, factorial, randomized block, and repeated measures. Is it a between (unpaired), within-subject (paired), or mixed design? Describe any counterbalancing required. Typical study designs for observation studies include cohort, cross sectional, and case-control studies.

8.2. Example: We have a between subjects design with 1 factor (sugar by mass) with 4 levels.

A mechanistic study. A two-group Randomised Controlled Trial (RCT) with allocation at the level of the individual. Participants were randomised to treatment with SSRI plus 'active' Cognitive Bias Modification (CBM) therapy or treatment with SSRI plus 'sham' CBM therapy.

9. Randomization (optional)

9.1. If you are doing a randomized study, how will you randomize, and at what level?

9.2. Example: We will use block randomization, where each participant will be randomly assigned to one of the four equally sized, predetermined blocks. The random number list used to create these four blocks will be created using the web applications available at http://random.org.

9.3. More info: Typical randomization techniques include: simple, block, stratified, and adaptive covariate randomization. If randomization is required for the study, the method should be specified here, not simply the source of random numbers.

Participants were randomly assigned to one of two groups: (1) active CBM therapy or (2) sham CBM therapy. Randomisation was by means of a computer-generated code, implemented by an

individual not involved in the recruitment process. Randomisation was minimised on gender, age (<35 years; ≥35 years), and depression severity (PHQ-9 score <17 vs. ≥17). The minimisation variables were important prognostic indicators and ensured a balance between the two groups.

Sampling Plan

In this section we'll ask you to describe how you plan to collect samples, as well as the number of samples you plan to collect and your rationale for this decision. Please keep in mind that the data described in this section should be the actual data used for analysis, so if you are using a subset of a larger dataset, please describe the subset that will actually be used in your study.

10. Existing data:

10.1. Preregistration is designed to make clear the distinction between confirmatory tests, specified prior to seeing the data, and exploratory analyses conducted after observing the data. Therefore, creating a research plan in which existing data will be used presents unique challenges. Please select the description that best describes your situation. Please do not hesitate to contact us if you have questions about how to answer this question (prereg@cos.io).

10.1.1. Registration prior to analysis of the data: As of the date of submission, the data exist and you have accessed it, though no analysis has been conducted related to the research plan (including calculation of summary statistics). A common situation for this scenario when a large dataset exists that is used for many different studies over time, or when a data set is randomly split into a sample for exploratory analyses, and the other section of data is reserved for later confirmatory data analysis.

11. Explanation of existing data (optional)

11.1. If you indicate that you will be using some data that already exist in this study, please describe the steps you have taken to assure that you are unaware of any patterns or summary statistics in the data. This may include an explanation of how access to the data has been limited, who has observed the data, or how you have avoided observing any analysis of the specific data you will use in your study.

11.2. Example: An appropriate instance of using existing data would be collecting a sample size much larger than is required for the study, using a small portion of it to conduct exploratory analysis, and then registering one particular analysis that showed promising results. After registration, conduct the specified analysis

on that part of the dataset that had not been investigated by the researcher up to that point.

11.3. More info: An appropriate instance of using existing data would be collecting a sample size much larger than is required for the study, using a small portion of it to conduct exploratory analysis, and then registering one particular analysis that showed promising results. After registration, conduct the specified analysis on that part of the dataset that had not been investigated by the researcher up to that point.

At present, prior to registration, no analysis has been performed on any of the MILESTONE behavioural or neuroimaging data other than to ensure data quality. No researcher who has access to the behavioural or imaging data collected has been informed of the treatment allocation of any participant, i.e., the researchers are blind to treatment allocation.

12. Data collection procedures:

12.1. Please describe the process by which you will collect your data. If you are using human subjects, this should include the population from which you obtain subjects, recruitment efforts, payment for participation, how subjects will be selected for eligibility from the initial pool (e.g. inclusion and exclusion rules), and your study timeline. For studies that donÍt include human subjects, include information about how you will collect samples, duration of data gathering efforts, source or location of samples, or batch numbers you will use.

12.2. Example: Participants will be recruited through advertisements at local pastry shops. Participants will be paid $10 for agreeing to participate (raised to $30 if our sample size is not reached within 15 days of beginning recruitment). Participants must be at least 18 years old and be able to eat the ingredients of the pastries.

12.3. More information: The answer to this question requires a specific set of instructions so that another person could repeat the data collection procedures and recreate the study population. Alternatively, if the study population would be unable to be reproduced because it relies on a specific set of circumstances unlikely to be recreated (e.g., a community of people from a specific time and location), the criteria and methods for creating the group and the rationale for this unique set of subjects should be clear.

<u>Patient recruitment</u>

The study is based at the University of Bristol. Patients were recruited from primary care (GP practices set in the surrounding areas of Bristol, North Somerset, South Gloucestershire, BANES, Gloucestershire and Somerset), as well as outside of the NHS through self-referrals including via in-person adverts (posters in public areas of Bristol and Cardiff) and via social media through an online study recruitment agency (Lindus Health). Patients completed an expression of interest form including a screening questionnaire and completed a screening call to assess initial eligibility. Potentially eligible participants then took part in an online Baseline Assessment with a researcher which included the following questionnaires:

- Revised Clinical Interview Schedule (CIS-R)(1): a detailed psychiatric instrument that will give an ICD10 diagnosis.
- Sociodemographic questions will include age, gender, employment status, educational qualifications, ethnicity, housing, and marital status.
- Patient Health Questionnaire (PHQ-9)(2) a brief measure of depressive symptoms.
- General Anxiety Disorder questionnaire (GAD-7)(3) a brief measure of anxiety.
- Quality of Life Enjoyment and Satisfaction Questionnaire (QLES)(4) a brief measure of life enjoyment and satisfaction.
- Snaith-Hamilton Pleasure Scale (SHAPS)(5) a measure of anhedonia.
- Reduced Morningness-Eveningness questionnaire (rMEQ)(6).

Eligible participants (eligibility criteria given below) then gave informed consent and were randomised into the study.

Participant eligibility criteria

Inclusion criteria:

- Aged 18-55 years.
- Have a NEW or FIRST episode of depression (defined as not prescribed an antidepressant in the previous 6 months).
- Prescribed a course of SSRI antidepressant medication (sertraline, citalopram, dapoxetine, escitalopram, fluoxetine, fluvoxamine, paroxetine, zimelidine and vortioxetine).
- Score > 10 on PHQ-9.

Exclusion criteria:

- Prescribed an antidepressant in the previous 6 months.
- Alcohol or substance dependency.
- Bipolar disorder.

- Schizophrenia/Psychosis.
- Dementia.
- Currently under psychiatric care (including those referred but not yet seen) for depression.
- Unable to access online CBM sessions (via PC, laptop, smartphone, tablet).
- Cannot complete questionnaires unaided or would require an interpreter.
- Are taking part in another trial involving a psychological/drug therapy.
- Have been receiving a course of high intensity psychological therapy for depression or anxiety in the last 6 months.
- Have a contra indication for fMRI scanning/imaging:
    - Significant hearing impairment (aids cannot be worn in the scanner).
    - Significant visual impairment that is not corrected by glasses/contact lenses. e.g., double vision or loss of vision in one eye, severe cataracts.
    - Metal objects in or around the body which cannot be removed (braces, pacemaker, metal fragments, hearing devices, accidents involving metal fragments).
    - History of established central nervous system disease or injury (e.g., cerebro-vascular disease, multiple sclerosis, Parkinson's disease, traumatic brain injury).
    - Epilepsy, type 1 diabetes (insulin pump or electronic device attached), or thermoregulatory problems, including Raynaud's disease.
    - Location sensitive tattoos to the head, neck, or genital area; (patients exceeding; tattoos covering >5% of the body; longest dimension>20cm; or multiple tattoos <20cm apart will be discussed with the radiographer).
    - Body Mass Index >35 kg/m^2.
    - Too physically unwell to tolerate a 30-minute fMRI scan, including musculo-skeletal disorders which make lying supine and still difficult.
    - Claustrophobia.
    - Pregnant or trying-to-become pregnant.
    - Participants were asked not to take recreational drugs for 72 hours prior to each test session and not to drink alcohol for 24 hours prior to each test session.

Assessments

2-week follow up

Follow-up assessments were conducted at 2 and 6 weeks after randomisation. At 2-week follow-up, participant completed four computerised online therapy sessions (Cognitive Bias Modification training for facial emotion processing) over the course of one week. Each therapy session took about 8-12 minutes. After completing the therapy sessions, participants attended an MRI scan at Cardiff University Brain Research Imaging Centre (CUBRIC). This in-person

appointment included a final session of online therapy, an online questionnaire including mood assessment (PHQ-9, GAD-7, QLES, SHAPS), description and practice of fMRI tasks before MRI scanning. During the MRI scan, participants underwent 1) a structural scan of their brain, 2) resting-state fMRI, 3) fMRI scanning whilst completing an implicit facial emotion processing task, 4) fMRI scanning whilst completing a working memory (n-back) task, and 5) fMRI scanning whilst completing a reward learning task. The appointment lasted roughly 2 hours in total. Participants were reimbursed £100 in Love2Shop vouchers plus travel expenses for completing this appointment.

6-week follow up

Participants completed a repeat assessment of all mood questionnaires online plus a final online therapy session. Participants were reimbursed a further £10 in Love2Shop vouchers for completing this follow up session.

Cognitive Bias Modification therapy

The active Cognitive Bias Modification (CBM) therapy consists of five computerised online therapy sessions in which participants are presented with brief psycho-educational information about the role of cognitive biases in depression, and then perform an intervention in which they judge the emotion displayed on screen in a simple, two alternative forced choice procedure. Each therapy session takes about 8-12 minutes to complete. The intervention targets the recognition of facial expressions of emotions in a set of images with ambiguous expressions (a 15-face continuum in which faces morph from happiness to sadness, see Figure 2). This therapy leads to an increased tendency to make positive attributions to ambiguous emotional stimuli (i.e. the cognitive bias that leads to negative interpretation of emotional faces is attenuated).

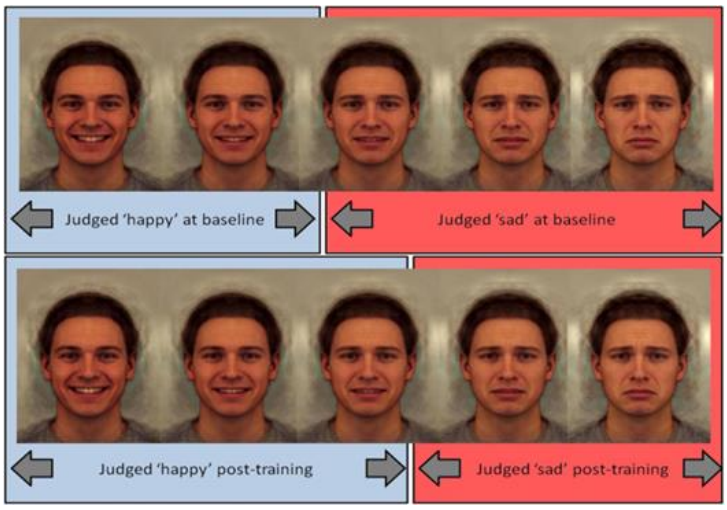


Figure 2. Training consists of feedback to shift the participant's balance point, estimated by presenting exemplar faces from a 15-frame morphed face image continuum using a two-alternative forced choice procedure. In the training condition, the 'correct' classification is shifted towards 'happy'; the two images nearest the balance point that the participant would have previously classified as 'sad' at baseline are considered 'happy' in terms of feedback. Feedback in the control condition is based directly on baseline performance and has no effect on responses. Sessions last 8 minutes and are fully automated.

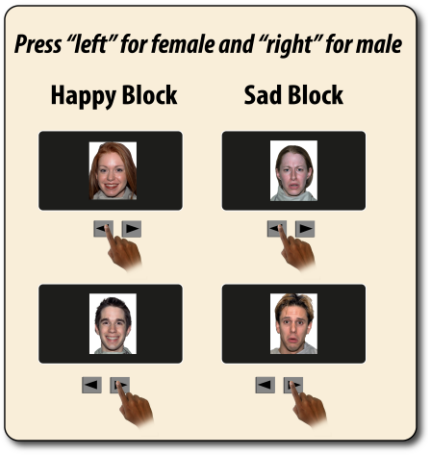
The CBM therapy consists of a baseline session, 5 training sessions (completed over 1 week and during 2-week follow up appointment), a repeat baseline session and a 6-week follow up repeat

baseline session. The baseline sessions consist of 45 trials (with each face of the continuum presented three times). The task requires the participant to make a forced choice judgement as to whether the presented face is displaying a 'sad' or 'happy' expression. Images are presented one at a time, in random order, for 150 ms. Stimuli are preceded by a fixation cross, which is presented for a random period ranging from 1500 to 2500 ms. Subsequent to presentation, and to prevent processing of afterimages, a backward mask of noise is presented for 250 ms, followed by a prompt asking the participant to respond. This remains on screen until the participant makes a response (i.e., a judgement of 'sad' or 'happy'). This baseline block allows us to determine where in the continuum the participant switches from making primarily 'happy' judgements to primarily 'sad' judgements.

Following the baseline block, training blocks are presented to participants. In the 'active' CBM therapy group, the feedback is tailored such that the two faces closest to the balance point that was initially categorised as sad by the participant are classed as 'incorrect', and it should be categorised as 'happy' (see fig 2). This training shifts individual balance point through feedback learning. In the 'sham' CBM therapy group, training blocks are the same except the feedback given is based directly on the participant's baseline threshold – so, the feedback does not attempt to change bias, as it reflects the participants response baseline. Our earlier work shows that this procedure is perceptually hard to distinguish from the active therapy, and that it has no effects on emotion perception. In the training blocks, there are 31 trials, in which images 1-2 (unambiguously happy) and 14-15 (unambiguously sad) are presented once, images 3-5 and 11-13 presented twice, and images 6-10 are presented three times. Different faces are presented on each day of training.

<u>fMRI Face Processing Task</u>

The task is modelled on a previous task (7) and was a simple blocked design face perception task involving the presentation of sad, happy and fearful facial expressions. In this fMRI task, thirteen 30 second blocks of a fixation cross (baseline condition) were interleaved with twelve 30 second



blocks of the emotional task (four blocks of sad faces (sad condition), four blocks of happy faces (happy condition) and four blocks of fearful faces (fearful condition). During each emotional block, participants viewed 10 emotional faces (5 female) presented on the screen one-by-one which were all derived from a standard set of pictures of facial affect (8). Each face was presented briefly (500 ms) and participants were asked to report the gender of the face (female/male) via an MRI compatible button box (see Figure 1). Within block interstimulus intervals (ISI) ranged between 2500 and 2900 ms (mean ISI 2900ms). To reduce potential carry-over effects, cycles of alternation between conditions were counterbalanced across participants. Stimulus presentation/participant button presses are registered and time-locked to fMRI data using E-Prime. Both accuracy (correct gender discrimination) and reaction times were recorded.

Figure 1: Face Processing Task This task involves the participant viewing faces with different emotional expressions. Participants are asked to judge the gender of the face (male or female).

MRI Data Acquisition

All data will be acquired using a 3T Siemens Prisma scanner (32-channel head coil) at the Cardiff University Brain Research Imaging Centre (CUBRIC). fMRI data will be acquired using a multiband (or simultaneous multi-slice) 2D echo-planar imaging (EPI) sequence (MB acceleration factor = 4, integrated parallel imaging technique (iPat) = OFF, Repetition time (TR) = 1500ms; TE = 32ms; 2.0mm x 2.0mm x 2.0mm voxels; Flip angle = 70°, anterior-to-posterior acquisition, interleaved slices, 519 volumes). A high-resolution structural scan was also acquired using a T1-weighted magnetization-prepared rapid gradient-echo (MPRAGE) sequence (TR = 2100ms; TE = 3.24ms; Inversion time = 850ms; flip angle = 8°; Field of View = 256mm x 256 mm^2 , 1.0 x 1.0 x 1.0 mm^3 voxel size, 176 slices) to aid with the preprocessing of the EPI data.

fMRI N-Back task

Participants are presented with a series of letters at the centre of the screen. Participants have to respond according to whether the currently presented letter on the screen matches or doesn't match the letter that was presented *n*-items. In this task, there are two conditions: 1-back and 2-back, as well as a fixation cross (rest) condition.

fMRI Reward Learning task

This task measures a participant's ability to fixate or change their behaviour in response to reward and punishment, respectively. Participants are able to choose between two items (shape objects). They are told that their goal is to find out, via trial and error, which item the computer has chosen as the correct answer. After selecting an item twice, they are presented with correct

or incorrect feedback that informs them whether they have, respectively, chosen the right or wrong shape. They need to continue to select the correct item until the answer changes (signaled by receiving 'incorrect' feedback) and switch responding to the other item.

13. Sample size (required)

- **13.1. Describe the sample size of your study. How many units will be analyzed in the study? This could be the number of people, birds, classrooms, plots, interactions, or countries included. If the units are not individuals, then describe the size requirements for each unit. If you are using a clustered or multilevel design, how many units are you collecting at each level of the analysis?**
- **13.2. Example: Our target sample size is 280 participants. We will attempt to recruit up to 320, assuming that not all will complete the total task.**
- **13.3. More information: For some studies, this will simply be the number of samples or the number of clusters. For others, this could be an expected range, minimum, or maximum number.**

Our target sample size is n = 72 (36 in each condition). We will recruit and randomise up to n = 84 participants assuming that ~20% attrition during the intervention period.

14. Sample size rationale (optional)

- **14.1. This could include a power analysis or an arbitrary constraint such as time, money, or personnel.**
- **14.2. Example: We used the software program G*Power to conduct a power analysis. Our goal was to obtain .95 power to detect a medium effect size of .25 at the standard .05 alpha error probability.**
- **14.3. More information: This gives you an opportunity to specifically state how the sample size will be determined. A wide range of possible answers is acceptable; remember that transparency is more important than principled justifications. If you state any reason for a sample size upfront, it is better than stating no reason and leaving the reader to "fill in the blanks." Acceptable rationales include: a power analysis, an arbitrary number of subjects, or a number based on time or monetary constraints.**

Our prior fMRI evidence (n = 36) indicates that our active CBM therapy intervention, relative to sham CBM therapy control, leads to increased amygdala activation in the happy > sad contrast in

the left amygdala (effect size of d = 0.69, p<0.05). A sample size of 72 will give >80% power to detect a similar effect in a replication sample at a 5% alpha level. We will recruit and randomise 84 participants (42 in each condition) into the study, allowing for ~20% attrition during the intervention period. Mood outcomes will be underpowered and therefore will be presented descriptively (i.e., without corresponding inferential statistics) but will provide estimates of likely effect size that may inform future trials of efficacy.

15. Stopping rule (optional)

15.1. If your data collection procedures do not give you full control over your exact sample size, specify how you will decide when to terminate your data collection.

15.2. Example: We will post participant sign-up slots by week on the preceding Friday night, with 20 spots posted per week. We will post 20 new slots each week if, on that Friday night, we are below 320 participants.

15.3. More information: You may specify a stopping rule based on p-values only in the specific case of sequential analyses with pre-specified checkpoints, alphas levels, and stopping rules. Unacceptable rationales include stopping based on p-values if checkpoints and stopping rules are not specified. If you have control over your sample size, then including a stopping rule is not necessary, though it must be clear in this question or a previous question how an exact sample size is attained.

N/A

Variables

In this section you can describe all variables (both manipulated and measured variables) that will later be used in your confirmatory analysis plan. In your analysis plan, you will have the opportunity to describe how each variable will be used. If you have variables which you are measuring for exploratory analyses, you are not required to list them, though you are permitted to do so.

16. Manipulated variables (optional)

16.1. Describe all variables you plan to manipulate and the levels or treatment arms of each variable. This is not applicable to any observational study.

16.2. Example: We manipulated the percentage of sugar by mass added to brownies. The four levels of this categorical variable are: 15%, 20%, 25%, or 40% cane sugar by mass.

16.3. More information: For any experimental manipulation, you should give a precise definition of each manipulated variable. This must include a precise description of the levels at which each variable will be set, or a specific definition for each categorical treatment. For example, "loud or quiet," should instead give either a precise decibel level or a means of recreating each level. 'Presence/absence' or 'positive/negative' is an acceptable description if the variable is precisely described.

In this Randomised Controlled Trial (RCT), participants were randomised to either the 'active' CBM therapy (received feedback to shift categorisation of emotional facial expressions) or the 'sham' CBM therapy (received feedback to reinforce participant's original categorisation thereby not creating a shift). The intervention comprised of five online therapy sessions (lasting 5-10 minutes each) starting at the 2-week follow-up stage but spread over one week. The sessions involved a baseline assessment of emotional face categorisation, 5 training sessions and a repeat baseline assessment. A repeat baseline assessment was also done at 6-week follow-up. Therefore, the manipulated variable in this study is whether participants received 'active' or 'sham' CBM therapy.

17. Measured variables (required)

17.1. Describe each variable that you will measure. This will include outcome measures, as well as any predictors or covariates that you will measure. You do not need to include any variables that you plan on collecting if they are not going to be included in the confirmatory analyses of this study.

17.2. Example: The single outcome variable will be the perceived tastiness of the single brownie each participant will eat. We will measure this by asking participants 'How much did you enjoy eating the brownie' (on a scale of 1-7, 1 being 'not at all', 7 being 'a great deal') and 'How good did the brownie taste' (on a scale of 1-7, 1 being 'very bad', 7 being 'very good').

17.3. More information: Observational studies and meta-analyses will include only measured variables. As with the previous questions, the answers here must be precise. For example, 'intelligence,' 'accuracy,' 'aggression,' and 'color' are too vague. Acceptable alternatives could be 'IQ as measured by Wechsler Adult Intelligence Scale' 'percent correct,' 'number of threat displays,' and 'percent reflectance at 400 nm.'

<u>Intervention check</u>

1) The post-training balance point at 2-week follow up and 6-week follow up for active and sham conditions (calculated as (number of 'happy' responses / 45 trials) x 15).

2) Effective learning rates for active and sham conditions (calculated using computational model, see below for information).

fMRI Behaviour

1) Gender identification accuracy (% correct classification of male/female faces) during fMRI face processing task. This is to ensure task engagement.

fMRI Neural Activation

Primary outcome: ROI activation in bilateral amygdalae in response to happy vs. sad faces.

Secondary outcomes:

1) ROI activation in bilateral amygdalae in response to happy vs. fear faces.
2) ROI activation in bilateral amygdala plus medial Prefrontal Cortex (mPFC) in response to happy vs. sad and happy vs. fear faces.

Exploratory outcomes:

1) ROI activation (bilateral amygdalae, mPFC) in response to sad vs. fear faces.
2) Whole-brain activation in response to happy vs. sad, happy vs. fear, and sad vs. fear faces.
3) Assessments of depressive symptoms (PHQ-9), anxiety symptoms (GAD-7), quality of life (QLES) and anhedonia (SHAPS) at baseline, 2-week and 6-week follow up.

18. Indices (optional)

18.1. If any measurements are going to be combined into an index (or even a mean), what measures will you use and how will they be combined? Include either a formula or a precise description of your method. If your are using a more complicated statistical method to combine measures (e.g. a factor analysis), you can note that here but describe the exact method in the analysis plan section.

18.2. **Example**: We will take the mean of the two questions above to create a single measure of 'brownie enjoyment.'

18.3. **More information**: If you are using multiple pieces of data to construct a single variable, how will this occur? Both the data that are included and the formula or weights for each measure must be specified. Standard summary statistics, such as "means" do not require a formula, though more complicated indices require either the exact formula or, if it is an established index in the field, the index must be unambiguously defined. For example, "biodiversity index" is too broad, whereas "Shannon's biodiversity index" is appropriate.

The CBM balance point will be calculated as: (number of 'happy' responses / number of trials in block)* 15.

An updated version of an ALCOVE computational model of associative learning (9) will be fitted to trial-by-trial responses during CBM training and will be used to compute the outcome measure; effective learning rate ($\varepsilon_{\text{effthr}}$). Briefly, the effective learning rate is a measure of how quickly the participant incorporates task feedback to update their responses to new faces. People generalize what they learn about each face stimulus according to its relationship to neighbouring stimuli on the facial expression morph continuum. This is accounted for in the effective learning rate which adjusts for individual training threshold. Higher effective learning rates indicate the participants respond more quickly and effectively to task feedback and therefore shift their classifications of ambiguous faces accordingly. Other parameters we will explore include: 1) inverse temperature (θ) (response reliability), 2) emotion bias (gH – gS) (positive values represent excess sad judgements on overt facial expressions), and 3) pre-training indifference point (a model-based measure analogous to balance point).

Analysis Plan

You may describe one or more confirmatory analysis in this preregistration. Please remember that all analyses specified below must be reported in the final article, and any additional analyses must be noted as exploratory or hypothesis generating.

A confirmatory analysis plan must state up front which variables are predictors (independent) and which are the outcomes (dependent), otherwise it is an exploratory analysis. You are allowed to describe any exploratory work here, but a clear confirmatory analysis is required.

19. Statistical models (required)

19.1. What statistical model will you use to test each hypothesis? Please include the type of model (e.g. ANOVA, multiple regression, SEM, etc) and the specification of the model (this includes each variable that will be included as predictors, outcomes, or covariates). Please specify any interactions, subgroup analyses, pairwise or complex contrasts, or follow-up tests from omnibus tests. If you plan on using any positive controls, negative controls, or manipulation checks you may mention that here. Remember that any test not included here must be noted as an exploratory test in your final article.

19.2. Example: We will use a one-way between subjects ANOVA to analyze our results. The manipulated, categorical independent variable is 'sugar' whereas the dependent variable is our taste index.

19.3. More information: This is perhaps the most important and most complicated question within the preregistration. As with all of the other questions, the key is to provide a specific recipe for analyzing the collected data. Ask yourself: is enough detail provided to run the same analysis again with the information provided by the user? Be aware for instances where the statistical models appear specific, but actually leave openings for the precise test. See the following examples:

19.3.1.1. If someone specifies a 2x3 ANOVA with both factors within subjects, there is still flexibility with the various types of ANOVAs that could be run. Either a repeated measures ANOVA (RMANOVA) or a multivariate ANOVA (MANOVA) could be used for that design, which are two different tests.

19.3.1.2. If you are going to perform a sequential analysis and check after 50, 100, and 150 samples, you must also specify the p-values you'll test against at those three points.

Intervention check

Data cleaning

Balance points that fall 3 times above or below the interquartile range will be considered outliers and will be removed. Data will be assessed for normality using skewness and kurtosis statistics and transformed if necessary.

Statistics

Linear regression models will be conducted to investigate the effectiveness of CBM training on the cognitive target (classification of ambiguous faces) where condition (active/sham CBM training) will be the predictor variable and post-training balance point at 2-week and 6-week follow-up will be the outcome variables. This model will be adjusted for the pre-training balance point , age, gender and depression severity (PHQ-9 score) at baseline.

fMRI Behaviour

Data cleaning

Participants that score below chance (<50% accuracy) in the gender discrimination task will be removed from analyses.

Statistics

A logistic regression will be conducted to investigate whether gender discrimination accuracy (outcome variable) is different between active vs. sham CBM training condition (predictor variable).

fMRI Neural Activation

Pre-processing

The fMRI data will be pre-processed and analysed using Statistical Parametric Mapping, version 12 (SPM 12, Wellcome Department of Cognitive Neurology, London, UK, https://www.fil.ion.ucl.ac.uk/spm/software/spm12). If necessary, the structural and functional images will first be manually reoriented so that the origin is reset over the anterior commissure. Then, the functional images will be unwarped (using field maps) and realigned to the first image by estimating six rigid-body realignment parameters (x, y, z, roll, pitch, yaw) in order to correct for the effects of head motion. The resulting images will then be slice-time corrected to the first slice in each acquisition to account for temporal disparities in slice time acquisition (to take into account the multiband sequence). Functional images will be co-registered to structural images and normalised to Montreal Neurological Institute (MNI) space using tissue probability maps and forward deformation fields from segmented structural images. and spatially smoothed using an 6mm Full Width Half Maximum (FWHM) Gaussian kernel.

First-level modelling

In this block design, the onsets and durations for three explanatory variables of interest will be modelled: 'happy faces', 'sad faces' and 'fearful faces'. These task-related regressors will be convolved with a canonical hemodynamic response function (hrf). Data will be filtered using a temporal high pass filter of 128 secs. Nuisance regressors will include 24 movement parameters (6 parameters estimated during realignment, 6 that are the square of these parameters, 6 that are the "spin-history" of realignment (realignment from preceding volume) and 6 that are the square of the spin-history parameters). Beta difference images will be created for the contrasts of interest: happy > sad, happy > fear, and sad > fear.

Second-level modelling

At the group level, individual first-level beta difference images representing the contrasts of interest will be submitted to separate mixed-effects modelling. For the primary analysis (H1), we will compare estimated beta values for viewing happy > sad faces between conditions

(active/sham CBM training) using an independent t-test (1, -1). For secondary analyses, we will perform similar t-tests comparing active and sham CBM training conditions for other constructed contrasts of interest (e.g., happy > fear, sad > fear). We will report unadjusted models, and models adjusted for age, gender, illness duration, antidepressant dose and duration.

Region of Interest analyses

Analyses will be primarily conducted using two *a priori* regions of interest (ROIs); 1) bilateral amygdalae, and 2) bilateral amygdalae plus medial prefrontal cortex (mPFC). The ROIs are selected based on an independent pilot dataset using this task (10) and literature emphasising the role of this network in processing of emotions (11). ROI masks for amygdala will be binary masks that will be anatomically defined using the probabilistic Harvard Oxford Subcortical and Cortical Structural atlases selecting voxels with >50% probability of lying within the structure. Left and right amygdalae will be merged into a single ROI mask. For the mPFC ROI, we will use the peak coordinates (MNI 8, 56, 18) evoked by psychotherapy which are reported in a recent meta-analysis of the neural effects of antidepression medication and psychological treatments (11). We will construct a 10mm sphere around the peak coordinates for the mPFC ROI.

For the primary and secondary outcomes, we will perform ROI analyses and significant effects will be reported if they survive small volume correction (SVC) with a voxel level threshold of $p < 0.05$ FWE-corrected (family-wise error).

Whole-brain analyses

Exploratory whole-brain analyses will also be conducted to test for any group differences in activity outside of the *a-priori* ROIs. Whole-brain analyses will be performed using a cluster defining threshold of $p < 0.001$ uncorrected, $k = 20$. We will report clusters surviving a FWE-corrected threshold of $p < 0.05$.

20. Transformations (optional)

20.1. If you plan on transforming, centering, recoding the data, or will require a coding scheme for categorical variables, please describe that process.

20.2. Example: The "Effect of sugar on brownie tastiness" does not require any additional transformations. However, if it were using a regression analysis and each level of sweet had been categorically described (e.g. not sweet, somewhat sweet, sweet, and very sweet), 'sweet' could be dummy coded with 'not sweet' as the reference category.

20.3. More information: If any categorical predictors are included in a regression, indicate how those variables will be coded (e.g. dummy coding, summation coding, etc.) and what the reference category will be.

N/A

21. Inference criteria (optional)

21.1. What criteria will you use to make inferences? Please describe the information youÍll use (e.g. p-values, bayes factors, specific model fit indices), as well as cut-off criterion, where appropriate. Will you be using one or two tailed tests for each of your analyses? If you are comparing multiple conditions or testing multiple hypotheses, will you account for this?

21.2. Example: We will use the standard p<.05 criteria for determining if the ANOVA and the post hoc test suggest that the results are significantly different from those expected if the null hypothesis were correct. The post-hoc Tukey-Kramer test adjusts for multiple comparisons.

21.3. More information: P-values, confidence intervals, and effect sizes are standard means for making an inference, and any level is acceptable, though some criteria must be specified in this or previous fields. Bayesian analyses should specify a Bayes factor or a credible interval. If you are selecting models, then how will you determine the relative quality of each? In regards to multiple comparisons, this is a question with few "wrong" answers. In other words, transparency is more important than any specific method of controlling the false discovery rate or false error rate. One may state an intention to report all tests conducted or one may conduct a specific correction procedure; either strategy is acceptable.

Computation of *p* values will be used to support group inferences. Two separate ROI analyses will be conducted and significant effects will be reported if they survive small volume correction (SVC) with a voxel level threshold of $p < 0.05$ FWE-corrected (family-wise error). Whole-brain analyses will be performed using a cluster defining threshold of $p < 0.001$ uncorrected, $k = 20$. We will report clusters surviving a FWE-corrected threshold of $p < 0.05$. Imaging data will be corrected for multiple comparisons using random field method.

22. Data exclusion (optional)

22.1. How will you determine what data or samples, if any, to exclude from your analyses? How will outliers be handled? Will you use any awareness check?

22.2. Example: No checks will be performed to determine eligibility for inclusion besides verification that each subject answered each of the three tastiness indices. Outliers will be included in the analysis.

22.3. More information: Any rule for excluding a particular set of data is acceptable. One may describe rules for excluding a participant or for identifying outlier data.

Any participants with incidental findings (e.g., cerebral malformation picked up during scan acquisition) or who were recruited in error (e.g., taking wrong medication) will be removed from the final analysis dataset. fMRI data quality will be checked for movement. Participants with problematic and unrepairable movement will be removed from the final analysis.

23. Missing data (optional)

23.1. How will you deal with incomplete or missing data?

23.2. Example: If a subject does not complete any of the three indices of tastiness, that subject will not be included in the analysis.

23.3. More information: Any relevant explanation is acceptable. As a final reminder, remember that the final analysis must follow the specified plan, and deviations must be either strongly justified or included as a separate, exploratory analysis.

Participants with missing imaging data will not be included in the final dataset.

24. Exploratory analysis (optional)

24.1. If you plan to explore your data set to look for unexpected differences or relationships, you may describe those tests here. An exploratory test is any test where a prediction is not made up front, or there are multiple possible tests that you are going to use. A statistically significant finding in an exploratory test is a great way to form a new confirmatory hypothesis, which could be registered at a later time.

24.2. Example: We expect that certain demographic traits may be related to taste preferences. Therefore, we will look for relationships between demographic variables (age, gender, income, and marital status) and the primary outcome measures of taste preferences.

We will conduct the following exploratory analyses:

1) We will perform two ROI analyses (bilateral amygdalae and bilateral amygdalae plus mPFC) to compare estimated beta values between conditions (active/sham CBM training) for sad > fear faces using an independent t-test (1, -1).
2) We will perform exploratory whole-brain analyses to test for any group differences in activity outside of the *a-priori* ROIs in response to happy > sad, happy > fear, and sad > fear.
3) To explore the effect of CBM training on neural activation, we will extract and plot individual beta values from peak activations showing significant group differences between active and sham conditions.
4) To understand the effect of different types of SSRI medication, we will conduct a sensitivity analysis on our primary analysis (group differences in amygdala activation in response to happy > sad faces) restricted only to patients taking Sertraline (most commonly prescribed SSRI).
5) To understand whether the effect of CBM training on neural activation is related to the length of time the patient has been on SSRI medication, we will extract individual beta values from peak activations showing significant group differences using the marsbar toolbox in SPM, and assess correlation with medication duration (split by active/sham condition).
6) Similarly, to understand the effect of age on CBM, we will: 1) examine the effect of age on CBM behaviour, and 2) examine the effect of age on neural activation by extracting individual beta values from peak activations showing significant group differences and assess correlation with age (split by active/sham condition).
7) We will test whether CBM training (active/sham predictor variables) is a significant predictor of depressive symptoms (PHQ-9), anxiety symptoms (GAD-7), quality of life (QLES) and anhedonia (SHAPS) at 2-week and 6-week follow up using a linear regression, adjusting for baseline measures, age and gender.
8) We will test whether peak amygdala activation (predictor variable) is a significant predictor of depressive symptoms (PHQ-9), anxiety symptoms (GAD-7), quality of life (QLES) and anhedonia (SHAPS) at 2-week and 6-week follow up using a linear regression, adjusting for baseline measures, age and sex.

Other

25. Other (Optional)

25.1. If there is any additional information that you feel needs to be included in your preregistration, please enter it here. Literature cited, disclosures of any related work such as replications or work that uses the same data, or other context that will be helpful for future readers would be appropriate here.

As this is a mechanistic study, patients that are randomised to one condition but complete the opposite condition (by experimenter error) will be treated as part of the condition they complete.

References

1. Lewis G, Pelosi A, Araya R, Dunn G. Measuring psychiatric disorder in the community: a standardized assessment for use by lay interviewers. Psychological medicine. 1992;22(2):465-86.
2. Kroenke K, Spitzer R, Williams J. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001;16(9):606-13.
3. Spitzer R, Kroenke K, Williams J, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med. 2006;166(10):1092-7.
4. Endicott J, Nee J, Harrison W, Blumenthal R. Quality of Life Enjoyment and Satisfaction Questionnaire: a new measure. Psychopharmacol Bull. 1993;29(2):321-6.
5. Snaith RP, Hamilton M, Morley S, Humayan A, Hargreaves D, Trigwell P. A scale for the assessment of hedonic tone the Snaith-Hamilton Pleasure Scale. Br J Psychiatry. 1995;167(1):99-103.
6. Adan A, . and Almirall, H. Horne & Östberg morningness-eveningness questionnaire: A reduced scale. Personality and Individual Differences. 1991;12(3):241-53.
7. Godlewska BR, Norbury R, Selvaraj S, Cowen PJ, Harmer CJ. Short-term SSRI treatment normalises amygdala hyperactivity in depressed patients. Psychological Medicine. 2012;42(12):2609-17.
8. Tottenham N, Tanaka JW, Leon AC, McCarry T, Nurse M, Hare TA, et al. The NimStim set of facial expressions: Judgments from untrained research participants. Psychiatry Research. 2009;168(3):242-9.
9. Stoddard J, Haller SP, Costa V, Brotman MA, Jones M. A Computational Model Reveals Learning Dynamics During Interpretation Bias Training With Clinical Applications. Biol Psychiatry Cogn Neurosci Neuroimaging. 2023.
10. Penton-Voak IS, Adams S, Button KS, Fluharty M, Dalili M, Browning M, et al. Emotional recognition training modifies neural response to emotional faces but does not improve mood in healthy volunteers with high levels of depressive symptoms. Psychological Medicine. 2021;51(7):1211-9.
11. Nord CL, Barrett LF, Lindquist KA, Ma Y, Marwood L, Satpute AB, et al. Neural effects of antidepressant medication and psychological treatments: a quantitative synthesis across three meta-analyses. Br J Psychiatry. 2021;219(4):546-50.