

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26

Multimodal neuroimaging data from a 5-week heart rate variability biofeedback randomized clinical trial

Hyun Joo Yoo¹, Kaoru Nashiro¹, Jungwon Min¹, Christine Cho¹, Noah Mercer¹, Shelby L. Bachman¹, Padideh Nasser¹, Shubir Dutt¹, Shai Porat¹, Paul Choi¹, Yong Zhang², Vardui Grigoryan³, Tiantian Feng¹, Julian F. Thayer⁴, Paul Lehrer⁵, Catie Chang⁶, Jeffrey A. Stanley⁷, Vasilis Z. Marmarelis¹, Shrikanth Narayanan¹, Daniel A. Nation⁴, Jessica Wisnowski¹, and Mara Mather^{1*}

¹University of Southern California

²University of Groningen

³University of California, Los Angeles

⁴University of California, Irvine

⁵Rutgers University

⁶Vanderbilt University

⁷Wayne State University School of Medicine

corresponding author(s): Mara Mather (mara.mather@usc.edu)

27 **Abstract**

28 We present data from the Heart Rate Variability and Emotion Regulation (HRV-ER) randomized
29 clinical trial testing effects of HRV biofeedback. Younger (N = 121, ages 18-31, 61 female) and
30 older (N = 72, ages 55-80, 45 female) participants completed baseline magnetic resonance
31 imaging (MRI) including T₁-weighted, functional MRI (fMRI), pulsed continuous arterial spin
32 labeling (PCASL), and proton magnetic resonance spectroscopy (¹H MRS). They also completed
33 an emotion regulation task during fMRI. During fMRI scans, physiological measures (blood
34 pressure, pulse, respiration, and end-tidal CO₂) were continuously acquired. Participants were
35 randomized to either increase heart rate oscillations or decrease heart rate oscillations during
36 daily sessions. After 5 weeks of HRV biofeedback, they repeated the baseline measurements
37 in addition to a few new measures (ultimatum game fMRI, training mimicking during blood
38 oxygen level dependent (BOLD) and PCASL fMRI). Participants also wore a wristband sensor to
39 estimate sleep time. Psychological assessment comprised three cognitive tests as well as ten
40 questionnaires related to emotional well-being. Data is publicly available via the OpenNeuro
41 data sharing platform.

42 **Background & Summary**

43 Heart rate variability (HRV) is one of the most consistent correlates of psychological and
44 emotional well-being and stress¹⁻³. However, it is not just random variation in the interval
45 between heartbeats that is associated with well-being. In healthy resting people, heart rate is
46 tonically suppressed by signals transmitted via the vagus nerve. This suppression of heart rate
47 is stronger when exhaling than when inhaling⁴, and it is “vagal HRV” or the high frequency
48 oscillations in heart rate in response to breathing that are most strongly associated with
49 positive well-being (or low negative affect or stress). Spending time every day breathing at a
50 pace of around 10 seconds per breath (a pace that induces resonance with the baroreflex and
51 so induces high oscillations in heart rate) while getting biofeedback on heart rate oscillatory
52 activity can enhance well-being^{5,6}. This suggests that heart rate oscillatory activity serves as
53 more than a readout of the integrity of the brain’s autonomic regulatory systems. Short bouts
54 of high heart rate oscillations may stimulate these regulatory systems, enhancing their
55 function⁷. To test this hypothesis, in a randomized clinical trial (ClinicalTrials.gov
56 NCT03458910), we scanned younger and older participants while at rest and while doing an
57 emotion regulation task both before and after five weeks of daily practice sessions in which
58 they received heart rate variability biofeedback to either increase or decrease heart rate
59 oscillations.

60 Initial studies using heart rate variability biofeedback yielded promising results and there has
61 been a significant growth in interest in this intervention^{5,6}. Compared with most prior HRV-
62 biofeedback studies, our study has a larger N and a more extensive set of outcome measures.
63 It is also unique among HRV-biofeedback studies in having all of the following features:
64 functional and structural brain outcome measures, a well-matched active comparison group,
65 inclusion of two age groups, and heart rate data available from each practice session. Thus,
66 these Heart Rate Variability and Emotion Regulation (HRV-ER) clinical trial data should be a
67 rich source for a variety of secondary analyses, including those investigating individual-
68 difference factors that affect responses to HRV-biofeedback, examination of age differences
69 in response to the intervention, and specific patterns of brain changes in response to the
70 intervention. Furthermore, the baseline pre-intervention data could be relevant for potential
71 secondary analyses unrelated to heart rate biofeedback. For instance, the larger N than seen
72 in most fMRI emotion regulation studies allows for individual difference comparisons,
73 especially given all the additional physiological data collected from each participant. In
74 addition, this study includes data not typically available in public datasets, such as PCASL, a
75 turbo spin echo (TSE) structural sequence targeting the locus coeruleus, and biochemical
76 measurements using proton magnetic resonance spectroscopy (¹H MRS), allowing for unique
77 secondary analyses not previously feasible.

78 **Methods**

79 **Power Considerations for sample size**

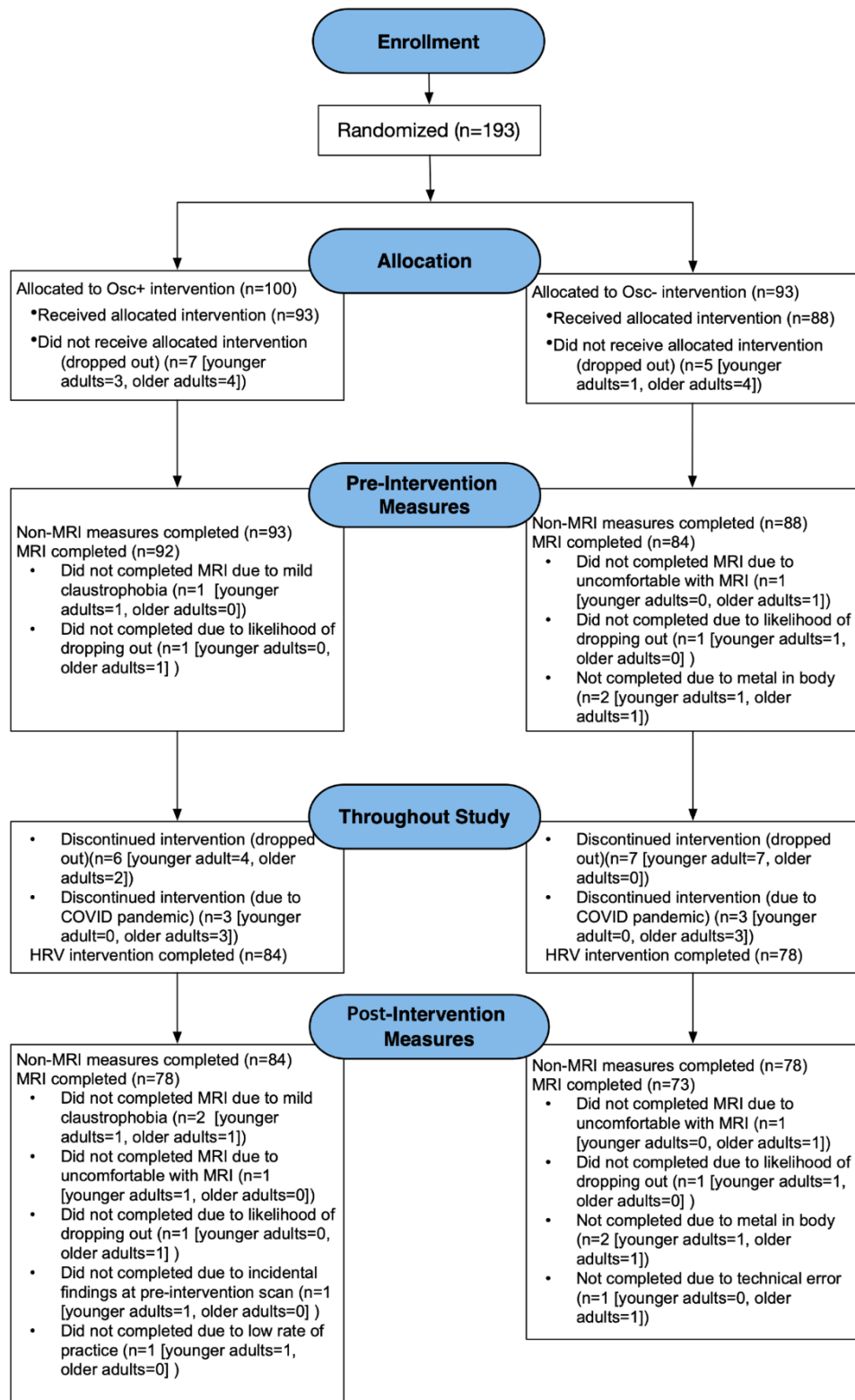
80 No prior studies had examined effects of these interventions on brain function so we were
81 unable to estimate effect sizes based on prior neuroimaging data. We elected to power our
82 study to detect medium or larger effect sizes. Our main planned statistical comparisons were
83 repeated-measures ANOVAs with within-between interactions. For these, a total sample size

84 of 46 would give 90% power to detect moderate effect sizes of $f = .25$ with $\alpha = .05$, given an
85 assumed correlation among the repeated measures of .5 (1). We also planned to examine
86 within-subject change within each of the conditions. A sample size of 44 in each group would
87 give 90% power to detect within-group change effect sizes of $d = .5$ in a two-tailed t-test with
88 $\alpha = .05$ (1). Thus, we aimed for an $N = 100$ completion rate across the two groups for each
89 age group to be able to accommodate potential exclusions for movement during imaging or
90 other data quality issues.

91

92 **Participants**

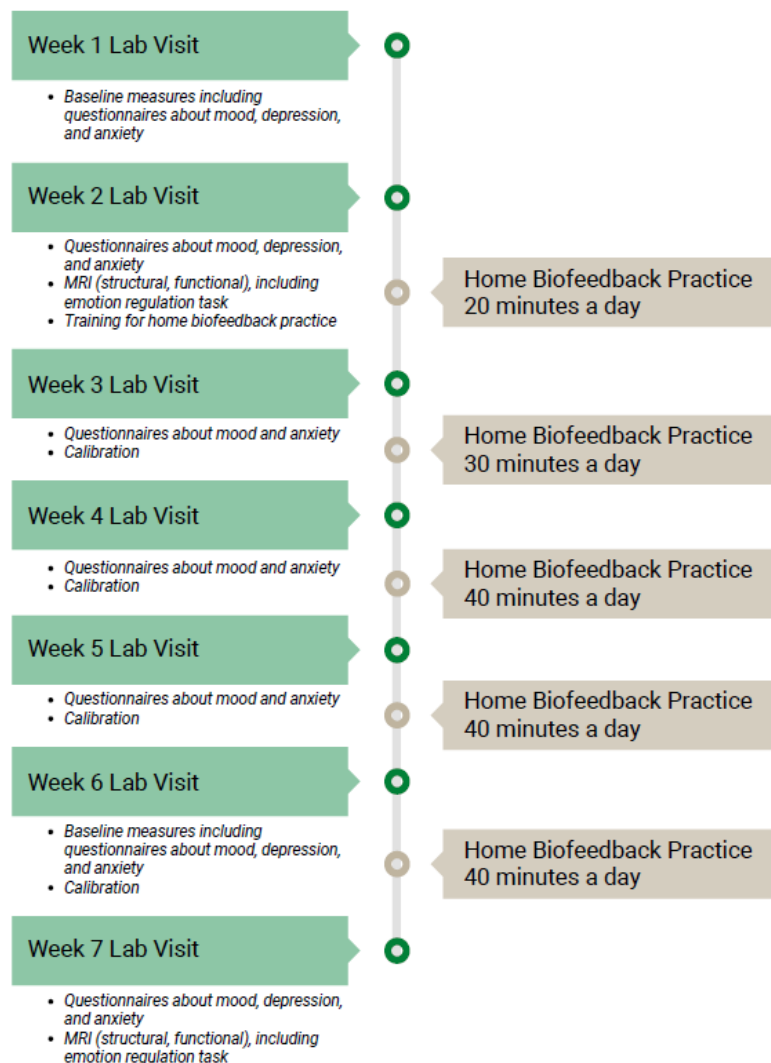
93 We recruited 121 younger participants aged between 18 and 35 years and 72 older
94 participants aged between 55 and 80 years via the USC Healthy Minds community subject
95 pool, a USC online bulletin board, Facebook and flyers between February 2018 and March
96 2020 (see Fig. 1 for drop-out rates per condition; note that older adult enrollment was cut
97 short by the COVID pandemic). Participants provided informed consent approved by the
98 University of Southern California (USC) Institutional Review Board. Participants were recruited
99 in waves of approximately 20 participants at a time, each of whom was assigned to a small
100 group of 3-6 people. Groups met for weekly lab visits at the same time and day each week.
101 After group assignments of a wave were complete, we assigned each group to one of two
102 conditions involving daily biofeedback that aimed to increase heart rate oscillations (Osc+
103 condition) or decrease heart rate oscillations (Osc- condition). We used blocked randomization
104 across small groups to maintain balanced numbers of each condition. We did this by
105 determining how many groups were assigned to a condition; for example, if 2 out of 5 groups
106 in a previous wave were assigned the Osc+ condition, then 3 out of 5 groups were assigned
107 that condition in the next wave. Then we randomly allocated those conditions to groups. One
108 research staff member who was blinded to participants and small group assignment generated
109 the random numbers and assigned the conditions to each small group. The study utilized a
110 single-blinded design; the consent document did not mention that there were two conditions
111 and participants in both conditions were told that we were interested in how training to
112 control heart rate might influence emotional health and the functions of brain regions involved
113 in emotion regulation. Upon completing the study, participants were paid for their
114 participation and received bonus payments based on their individual and group performances
115 (incentives for training were the same across conditions; see Supplementary Information for
116 more details). Prospective participants were screened and excluded for major medical,
117 neurological, or psychiatric illnesses. We excluded people who had a disorder that would
118 impede performing the HRV biofeedback procedures (e.g., coronary artery disease, angina,
119 cardiac pacemaker), who currently were training using a relaxation, biofeedback or breathing
120 practice, or were on any psychoactive drugs other than antidepressants or anti-anxiety
121 medications. We included people who were taking antidepressant or anti-anxiety medication
122 and/or attending psychotherapy only if the treatment had been ongoing and unchanged for
123 at least three months and no changes were anticipated. Gender, education, age, and race did
124 not differ significantly in the two conditions.



125 **Fig. 1. CONSORT flow diagram showing participant flow through each stage of the**
 126 **randomized controlled trial.**

7- WEEK STUDY SCHEDULE

Lab visits and biofeedback training



127 **Fig. 2. Study 7-week schedule.**

128

129 **Overview of 7-week protocol schedule**

130 The study protocol involved seven weekly lab visits and five weeks of home biofeedback
131 training (Fig. 2). Each lab visit began with questionnaires assessing mood and anxiety. The first
132 lab visit involved non-MR baseline measurements, including several questionnaires. The
133 second lab visit involved the baseline MR session, followed by the first biofeedback calibration
134 and training session (see below for details). The weekly lab visits (except for weeks with MR
135 sessions) were run in small groups in which participants shared their experiences and tips
136 about biofeedback training with other participants from the same condition, while 1-2
137 researchers facilitated the discussion. Outside the lab, participants used a customized social
138 app to communicate with other group members and researchers about their progress on daily
139 biofeedback training. The Week 6 lab visit repeated the assessments from the first lab visit.
140 The final (7th) lab visit repeated the baseline MR session scans in the same order. Additional

141 training-session scans were collected at the end of the scan protocol. Finally, after the scan,
142 participants completed a post-study questionnaire. Table 1 provides detailed information
143 about the measurement at each time point.

144 **Table 1. Measurement at each time point.**

Category	Data type	Week1	Week2	Week3	Week4	Week5	Week6	Week7	Measurement
Demographics	Demographics	*							•Self-report
Questionnaire	Emotion-weekly (POMS, SAI)	*	*	*	*	*	*	*	•Self-report
	Emotion-4 times (CESD, TAI)	*	*				*	*	•Self-report
	Altruism		*						•Self-report
	others-pre/post (FFMQ)	*					*		•Self-report
Cognitive task	NIH toolbox: Cognition	*					*		•score
	SART	*					*		•score •response time
	Picture memory task				*	*			•encoding (week 4) •recognition (week 5) •recall (week 5)
Heart rate data	calibration		*	*	*	*	*	*	•Inter-beat-interval data from pulse
	Home training		*	*	*	*	*		•inter-beat-interval data from pulse
Stress task	Physiological data	*					*		•ECG, respiration, GSR, continuous blood pressure
	Behavioral data	*					*		•score
MR Imaging	Functional-resting		*					*	•brain imaging data •physiological data
	Functional-resting ASL		*					*	•brain imaging data •physiological data
	Functional-Emotion regulation		*					*	•brain imaging data •physiological data •event data
	Anatomical-T ₁		*					*	•brain imaging data
	¹ H MRS		*					*	•brain biochemistry data
	Anatomical-TSE		*					*	•brain imaging data (locus coeruleus)
	Functional-UG							*	•brain imaging data •physiological data •event data
	Functional-training mimicking							*	•brain imaging data •physiological data •event data
	Functional-training mimicking ASL							*	•brain imaging data •physiological data •event data

145

146 Biofeedback training

147 *Osc+ group*

148 Participants wore an ear sensor to measure their pulse. They viewed real-time heart rate
149 biofeedback while breathing in through the nose and out through the mouth in synchrony with
150 a visual pacer. The software⁸ provided a summary ‘coherence’ score for participants calculated
151 as peak power/(total power – peak power). Peak power was determined by finding the highest
152 peak within the range of 0.04 – 0.26 Hz and calculating the integral of the window 0.015 Hz
153 above and below this highest peak. Total power was computed for the 0.0033 – 0.4 Hz range.

154 During the second lab visit, we introduced participants to the device and had them complete
155 five minutes of paced breathing at 9 s, 10 s, 11 s, 12 s, and 13 s per breath, which approximately
156 corresponds with 6.5, 6, 5.5, 5 and 4.5 breaths/min in Lehrer et al.⁹ Next, we computed various
157 aspects of the oscillatory dynamics for each breathing pace using Kubios HRV Premium 3.1
158 software¹⁰ and assessed which breathing pace had the most of the following characteristics:
159 highest LF power, the highest maximum LF amplitude peak on the spectral graph, highest
160 peak-to-trough amplitude, cleanest and highest-amplitude LF peak, highest coherence score
161 and highest root mean squared successive differences (RMSSD). Participants were then
162 instructed to train at home with the pacer set to the frequency that appeared to best
163 approximate their resonance frequency⁹ and to maximize their coherence scores.

164 During the third visit, they were asked to complete three 5-min paced breathing segments:
165 the best condition from the last week’s visit, half a breath per minute faster, and half a breath
166 slower than the best condition. They were then instructed to train the following week at the
167 pace that appeared most likely to be a resonance frequency based on the characteristics listed
168 above. In subsequent weekly visits, during 5-min training segments, they were asked to try out
169 abdominal breathing and inhaling through the nose/exhaling through pursed lips as well as
170 other strategies of their choice.

171 *Osc- group*

172 This condition was designed to be as similar as possible to the Osc+ condition, but with the
173 opposing goal (to reduce heart rate oscillations). The same biofeedback ear sensor device was
174 used in this condition and participants were asked to practice for the same amount of time.
175 However, we created custom software to display a different set of feedback to the Osc-
176 participants¹¹. During each Osc- training session, a ‘calmness’ score was provided as feedback
177 to the participants instead of the coherence score. The calmness score was calculated by
178 multiplying the coherence score that would have been displayed in the Osc+ condition by -1
179 and adding 10 (an ‘anti-coherence’ score). The net result was that participants got more
180 positive feedback (higher calmness scores) when their heart rate oscillatory activity in the 0.04
181 – 0.26 Hz range was low. The software also included a minor point penalty when heart rate
182 was the lowest it had been in the most recent 15 s. Specifically, every 5 s, a local maximum IBI
183 was set based on the maximum IBI from the past 15 s. If, at that point, the participant’s current
184 IBI was longer than this local maximum, the calmness score displayed for the next 5 s was the
185 anti-coherence score - 2. Naturally, most of the time, current IBI was lower than the local
186 maximum, and in those cases, the calmness score was the anti-coherence score +1. Thus, there

187 was a penalty in their calmness score for moments when their heart rate was slower than it
188 had been in any of the past 15 s.

189 During the initial calibration session at the end of the second lab visit, each participant was
190 introduced to the device and feedback and was asked to come up with five strategies to lower
191 heart rate and heart rate oscillations. The participant was asked to wear the ear sensor and
192 view real-time heart rate biofeedback while they tried each strategy for five minutes. We
193 analyzed the data in Kubios and identified the best strategy as the one with the most of the
194 following characteristics: lowest LF power, the minimum LF-amplitude peak on the spectral
195 graph, lowest peak trough amplitude, multiple and lowest-amplitude LF-peak, highest
196 calmness score and lowest RMSSD. Participants were then instructed to use this strategy to
197 try to maximize their calmness scores in their home training sessions.

198 On the third visit, they were asked to select three strategies and try each out in a 5-min session.
199 The strategy identified as best (based on the same characteristics used in the initial calibration
200 session) was selected as the one to focus on during home training. In subsequent weekly visits,
201 during 5-min training segments, they were again asked to try out strategies of their choice.

202 Heart rate data from lab calibration and home training

203 During lab calibration, pulse was measured using HeartMath emWave pro software with an
204 infrared pulse plethysmograph (PPG) ear sensor while participants sat in a chair with knees at
205 a 90 degrees angle and both feet flat on the floor. The pulse wave was recorded with a
206 sampling rate of 370 Hz. Interbeat interval (IBI) data was extracted after eliminating ectopic
207 beats and other artifactual signals through a built-in process in emWave pro. On each home
208 training session, pulse was measured using the same devices and software used for calibration
209 sessions. An IBI data file was saved on a study-provided laptop and transferred to the lab
210 server after completing each session.

211 **Questionnaires**

212 Emotion questionnaires

213 During each lab visit, participants completed the profile of mood states (POMS)¹² and the State
214 Anxiety Inventory (SAI)¹³. We used the 40-item version of POMS. Participants reported how
215 much each item reflected how they felt at the moment using a scale from 1 (not at all) to 5
216 (extremely). Total mood disturbance was calculated by subtracting positive-item totals from
217 negative-item totals. A constant value (i.e. 100) was added to the total mood disturbance to
218 eliminate negative scores. The SAI measures state anxiety using 20 statements. Participants
219 indicated how they felt at the moment on a scale from 1 (not at all) to 4 (very much so). In
220 Weeks 1, 2, 6 and 7, we also administered the Trait Anxiety Inventory (TAI)¹³ and the Center
221 for Epidemiological Studies Depression Scale (CES-D) in Weeks 1, 2, 6 and 7¹⁴. The TAI
222 measures trait anxiety using 20 statements, which participants rated on a 4-point scale from
223 1 (not at all) to 4 (very much so). The CES-D consists of 20 statements, which participants rated
224 on a 4-point scale from 0 (rarely) to 3 (most or all of the time). At Week 1 and Week 6 lab visits,
225 participants completed six additional emotion questionnaires. We assessed trait mindfulness
226 using the 20-item version of the Five Facet Mindfulness Questionnaire (FFMQ)¹⁵. Participants
227 rated each item using a scale from 1 (never) to 5 (very often/always true). We also

228 administered the Smith Relaxation States Inventory 3 (SRSI3)¹⁶ to assess various aspects of
229 stress, relaxation, meditation, and mindfulness. Participants completed state and disposition
230 versions of the SRSI3 each consisting of 38 items. The state version asks how you “feel right
231 now” on a 6-point scale from 1 (not at all) to 6 (maximum). The disposition version asks how
232 often you experience relaxation states and stress states. We slightly modified the
233 disposition version and asked how often each item has been experienced “in the past month”
234 on a 6-point scale from 1 (rarely or never, less than once a month) to 6 (a lot, more than once
235 a day). We calculated SRSI3 scores based on the Relaxation/Meditation/Mindfulness (RMM)
236 Tracker/SRSI3 Manual v9.9.2020, which includes 34 items for scoring¹⁷. We measured the
237 extent and severity of fatigue using the 11-item Chalder Fatigue Scale (CFQ 11)¹⁸ on a 4-point
238 scale from 0 (less than usual) to 3 (much more than usual) or 0 (better than usual) to 3 (much
239 worse than usual). We also administered the 10-item Emotion Regulation Questionnaire
240 (ERQ)¹⁹, which is designed to measure tendency to regulate emotions in two ways (cognitive
241 reappraisal and expressive suppression) on a 7-point scale from 1 (strongly disagree) to 7
242 (strongly agree). Participants also completed a self-efficacy version of the 10-item ERQ, which
243 asks how “capable” they are of regulating their emotions on the same 7-point scale. In addition,
244 perceived stress was measured using the NIH Toolbox Perceived Stress Survey²⁰, a 10-item
245 version of the Perceived Stress Scale²¹. Participants rated the frequency of stressful
246 experiences and the extent to which they felt strained or overloaded during the past month
247 (e.g., How often have you felt nervous and “stressed”? How often have you felt difficulties
248 were piling up so high that you could not overcome them?) on a five-point scale, ranging from
249 Never (1) to Very Often (5); higher scores correspond to greater perceived stress. Score on the
250 perceived stress scale has been calculated as the mean score. All the rest are calculated as
251 sum scores.

252

253 Altruism questionnaire

254 Participants completed the Altruism scale questionnaire²² during their lab visits on the second
255 week of intervention. Altruism scale questionnaire is a self-report scale with 20 items each
256 describing an altruistic behavior (e.g., “I have done volunteer work for a charity” and “I have
257 delayed an elevator and held the door open for a stranger). Participants were instructed to
258 rate the frequency of engaging in these behaviors on a 5-point scale (1 = never; 2 = once,
259 3 = more than once; 4 = often; 5 = very often). Higher scores in this scale correspond with
260 higher altruistic tendencies.

261 Demographics and post-study questionnaire

262 The week-1 visit, participants completed questionnaires including basic demographics, clinical
263 history including medications.

264 After the Week-7 post-intervention scan, participants completed a questionnaire surveying
265 their experience during the study. They provided self-ratings of difficulty of daily heart rate
266 biofeedback training, level of effort to complete the training, expectations of the training
267 impact on well-being, and likelihood of continuing the training after the study’s conclusion.

268 **Stress task**

269 During the week 1 and week 6 lab visits, participants completed a task designed to assess
270 reactivity to and recovery from acute stress. The task consisted of several phases: a 4-minute
271 baseline resting phase, a stress phase, and a 4-minute recovery resting phase. Younger
272 participants completed two computerized tasks during the stress phase: a Paced Auditory
273 Serial Addition Task (PASAT) and a modified Stroop color-word matching task. Seven older
274 participants completed a pilot version of the stress phase which included both the PASAT and
275 Stroop tasks. But the rest of older participants completed only the Stroop task. Descriptions
276 of the PASAT and Stroop tasks are provided below. To enhance the socially evaluative nature
277 of the tasks, participants were told that their performance would be evaluated by the
278 experimenter and compared with that of other participants. In addition, auditory feedback
279 was provided during each task, with a buzzer sound played for incorrect or missed responses
280 and a bell sound played for correct responses²³. During the baseline and recovery resting
281 phases, participants sat with their feet resting flat on the ground, their hands resting in a prone
282 position on a flat table, and their eyes open.

283 Physiological signals were recorded during all phases of the stress task using a BIOPAC MP160
284 system at a sampling frequency of 2 KHz. Electrocardiogram (ECG) and respiration signals were
285 sent to the MP160 with a BioNomadix wireless transmitter. ECG signals were measured using
286 a Lead II configuration and disposable, pre-gelled Ag/AgCl electrodes (EL501). Respiration
287 signals were assessed with the BIOPAC Respiratory Effort Transducer, which involved a belt
288 being placed around the lower rib cage to measure changes in chest circumference. Blood
289 pressure signals were measured from the non-dominant arm with a BIOPAC noninvasive blood
290 pressure monitoring system (NIBP100D). For electrodermal activity recordings, disposable,
291 pre-gelled Ag/AgCl electrodes (EL507) were attached to the palmar side of the medial
292 phalange of the fourth and fifth fingers of each participant's non-dominant hand (as the index
293 and middle fingers were used for continuous blood pressure recordings); leads were
294 connected to a BIOPAC GSR100C module. Raw physiological signals were split into segments
295 corresponding to the phases of the stress task.

296 Paced Auditory Serial Addition Task (PASAT)

297 During the PASAT, participants were shown a series of digits. They were instructed to add
298 each digit to the digit shown previously and enter the sum using the keyboard²⁴. The sum
299 always ranged between 1-20. The PASAT consisted of 30 trials, and participants completed 4
300 practice trials prior to beginning the task.

301 Modified Stroop Task

302 On each trial of the modified Stroop color-word matching task, participants first saw one of
303 two instructions: "color" or "meaning," followed by a color word (RED, GREEN, or BLUE)
304 presented in one of three colors incongruent with its meaning (either red, green or blue)²⁵.
305 Participants were instructed to press a key corresponding to either the color or meaning of
306 the word stimulus, depending on the instruction shown immediately before the word. The
307 task consisted of 20 trials, and participants completed 4 practice trials before beginning the
308 task.

309 **Cognitive tasks**

310 The National Institutes of Health Toolbox Cognitive Battery

311 The National Institutes of Health (NIH) Toolbox Cognitive Battery is a component of the NIH
312 Toolbox for Assessment of Neurological and Behavioral Function (www.nihtoolbox.org;^{26,27})
313 that comprises extensively validated computer-administered cognitive tests for use across
314 childhood and adolescence, early adulthood, and old age. We administered the NIH-Toolbox
315 cognitive battery using an iPad app on a 9.7 inch iPad Air 2. The Flanker Test, the List Sorting
316 Working Memory (LSWM) Test, and the Pattern Comparison Processing Speed (PCPS) Test
317 were administered to evaluate attention and executive function, working memory, and
318 processing speed, respectively.

319 In addition to raw scores and/or reaction time, the NIH-toolbox cognitive battery generates
320 computed scores and three types of standard scores for each subtest: uncorrected standard
321 scores, age-corrected standard scores, and fully adjusted scores that account for age,
322 education, gender, and race/ethnicity.

323 *The Flanker inhibitory control and attention test*

324 The Flanker test is a version of the Eriksen flanker task derived from the Attention Network
325 Test²⁸. On each trial, a central directional target (arrows for ages 8 and older) is flanked by
326 similar stimuli on the left and right. The participant chooses the direction of the central
327 stimulus. On 12 congruent trials, the flankers face the same direction as the target. On 8
328 incongruent trials, they face the opposite direction. A scoring algorithm integrates accuracy
329 and reaction time, a measure more relevant to adult performance on this task, yielding
330 computed scores from 0 to 10. There are 20 trials for ages 8 and older, and the test duration
331 is about 4 min. Score is based on an algorithm derived from both accuracy and reaction time
332 if the former is >80%. If less than 80%, score is accuracy. The test takes approximately
333 3 minutes to administer.

334 *The List Sorting Working Memory Test*

335 This task is an adaptation of Mungas' List Sorting task from the Spanish and English
336 Neuropsychological Assessment Scales^{29,30}. In this task, a list of stimuli is presented both
337 visually (picture) and auditorily (recording of a one-word description of the stimulus) on a
338 computer monitor, one at a time at a rate of 2 sec per stimulus, and participants are required
339 to repeat all the stimuli back to the examiner in order of increasing real-world size, from
340 smallest to largest. In the first phase of the test (i.e., the 1-List phase), participants are first
341 shown a list with two items drawn from a single category (i.e., food). If participants are correct
342 on this 2-item list, the number of items in the list presented on the next trial increases by one
343 item, up to a total of seven items per list (i.e., list length ranges from a 2-item list to a 7-item
344 list, for a total of six levels of list length). If participants err on a trial at a given list length, they
345 receive another trial with the same number of items in the list; if they err on that trial, this
346 phase of the test is discontinued. Following the 1-list phase, all participants proceed to the
347 second phase of the test (the 2-list phase), in which they see lists of items drawn from two
348 different categories (i.e., food and animals). Participants are instructed to reorder and repeat
349 the stimuli first from one category, then the other, in order of size within each category. Lists

350 in the 2-list phase start with a 2-item list and increase in number of items in the same way as
351 in the 1-list phase. For both phases, for each list length, participants receive a score of 2 points
352 if they are correct on the first trial. A second trial at a given list length is only administered
353 when participants fail the first trial. Participants receive a score of 1 point only for a given list
354 length if they fail the first trial at that list length but pass the second trial. Test scores consist
355 of combined total trials correct on the 1-list and 2-list phases of the task. The test takes
356 approximately 7 minutes to administer.

357 *The Pattern Comparison Processing Speed Test*

358 The Pattern Comparison Processing Speed Test is modeled after Salthouse's Pattern
359 Comparison Task³¹. This test requires participants to identify whether two visual patterns are
360 the "same" or "not the same" (responses were made by pressing a "yes" or "no" button).
361 Patterns are either identical or vary on one of three dimensions: color, adding/taking
362 something away, or one versus many. Scores reflect the number of correct items (of a possible
363 130) completed in 90 s. The test takes approximately 3 minutes to administer.

364 Sustained Attention to Response Task (SART)

365 The SART³² was administered during Week 1 and Week 6 lab visits. During the task,
366 participants were presented with a random series of single-digit numbers, ranging from 1 to
367 9. Participants were instructed to press the spacebar as soon as they saw each number other
368 than 3. The task consisted of 225 trials where a single digit was presented for 250 ms with a
369 900-ms-lasting mask image between trials. The task took about 6 minutes and was based on
370 the web-based Inquisit SART task developed by Millisecond Software.

371 Picture memory tasks

372 The emotional memory task was administered at the Week 4 and Week 5 lab visits. Seventy-
373 two stimuli were selected from The Nencki Affective Picture System (NAPS)³³, a database of
374 realistic photographs that aim to induce positive, negative, or neutral emotional states. Stimuli
375 were first counterbalanced by valence (24 each of positive, negative, and neutral); then two
376 sets of 36 stimuli were created and counterbalanced by valence in each set (12 each of positive,
377 negative, and neutral). Participants completed the task on the Qualtrics Survey platform.

378 At the Week 4 visit, one of the two sets of 36 stimuli were shown to participants by random
379 assignment. After viewing each image, participants rated on a 1-9 scale how positive, negative,
380 or neutral they found the images (1 = very negative to 9 = very positive, 5 = neutral). No time
381 limit was imposed for participants to provide their ratings, and each image was shown on the
382 screen for 4 seconds. Once all images in the set were rated, participants completed a free
383 recall task, where they were asked to describe as many images they saw with as much detail
384 as possible. There was no time limit imposed for the free recall.

385 At the Week 5 visit, participants first completed the same free recall task as in Week 4.
386 Participants then viewed all 72 stimuli and for each image selected one of 3 response options:
387 Remember, Know, and New. "Remember" was described as having a vivid memory of an image
388 (i.e., it evoked thoughts or feelings when it was seen, recollections of something else that
389 happened at that same moment, or where in sequence the image was). "Know" was described

390 as being confident that the image was seen but that nothing specific related to thoughts,
391 feelings, or experience was associated with the image. “New” was described as being
392 confident that the image was not seen before. Each image was shown on the screen for 3
393 seconds, and there was no time limit for providing responses.

394 **Sleep time**

395 Sleep and HRV derived from slow-wave sleep were measured with WHOOP wristbands³⁴.
396 Participants were provided WHOOP wristbands on the first day until the final week of the
397 study. All participants were instructed to wear the wristband as close to 24 hours per day as
398 possible. Sleep data metrics include, but are not limited to daily information of: hours in bed,
399 hours of sleep, hours awake, hours of REM sleep, hours of deep sleep, hours of light sleep,
400 number of disturbances, number of sleep cycles, hours of naps. Average heart rate per minute,
401 resting heart rate, and heart rate variability were provided by WHOOP each day as well.
402 WHOOP calculates sleep-derived HRV during the final 5 minutes of recorded deep sleep.
403 Relative to gold standard polysomnography, WHOOP algorithms have been validated by
404 independent researchers as having a 95% sensitivity for sleep, 68% sensitivity for deep sleep
405 and 70% for REM sleep³⁵.

406 **MRI/MRS data acquisition**

407 MRI scan session order

408 In both the pre- and post-intervention MR sessions, scans were conducted in the following
409 order: 1) resting-state during BOLD fMRI; 2) resting-state during pCASL; 3) emotion regulation
410 task during fMRI; 4) T₁-weighted structural scan; 5) magnetic resonance spectroscopy (MRS);
411 and 6) T₁-weighted TSE scan. The post-intervention session included three additional scans,
412 which were performed between the ¹H MRS and TSE scans in the following order: 1) ultimatum
413 game task; 2) training-mimicking session during BOLD fMRI; and 3) training-mimicking session
414 during pCASL. During both training-mimicking scans, participants engaged in their daily
415 training without biofeedback (see below for details).

416 MRI scan parameters

417 We employed a 3T Siemens MAGNETOM Trio scanner with a 32-channel head coil at the USC
418 Dana and David Dornsife Neuroimaging Center. T₁-weighted 3D structural MRI brain scans
419 were acquired pre and post intervention using a magnetization prepared rapid acquisition
420 gradient echo (MPRAGE) sequence with TR = 2300 ms, TE = 2.26 ms, slice thickness = 1.0 mm,
421 flip angle = 9°, field of view = 256 mm, and voxel size = 1.0 x 1.0 x 1.0 mm³, with 175 volumes
422 collected (4:44 min). Functional MRI scans during resting-state, emotion-regulation, training
423 and ultimatum game tasks were acquired using multi-echo echo-planar imaging sequence with
424 TR= 2400 ms, TE 18/35/53 ms, slice thickness = 3.0 mm, flip angle = 75°, field of view = 240
425 mm, voxel size = 3.0 x 3.0 x 3.0 mm. We acquired 175 volumes (7:00 min) for the resting-state
426 scan and training scan, 250 volumes (10:00 min) for the emotion-regulation task and 244
427 volumes (9:45 min) for the ultimatum game task. PCASL scans were acquired with TR = 3880
428 ms, TE = 36.48 ms, slice thickness = 3.0 mm, flip angle = 120°, field of view = 240 mm and voxel
429 size = 2.5 X 2.5 X 3.0 mm³, with 12 volumes collected (3:14 min; 1st volume was an M0 image,
430 2nd volume was a dummy image that was discarded, and the remaining 10 volumes were five

431 tag-control pairs) both during resting-state (pre and post intervention) and training-mimicking
432 (post intervention) scans. This ASL approach provides high precision and signal-to-noise
433 properties and has better test-retest reliability than pulsed or continuous ASL techniques³⁶.
434 The two-dimensional, multi-slice TSE scan was acquired with TR = 750 ms; TE = 12 ms; flip
435 angle = 120°; bandwidth = 287 Hz/pixel; voxel size = 0.43 x 0.43 x 2.5 mm³, gap between slices
436 = 1.0mm, 11 axial slices). The ¹H MRS data were acquired using a single-voxel point-resolved
437 spectroscopy (PRESS) sequence with an echo time of 35 ms and repetition times of 2.0 s from
438 a 4.1 cm³ (1.6 x 1.6 x 1.6 cm³) voxel localized to the anterior portion of the anterior cingulate
439 cortex. Axial, sagittal, and coronal orientations were assessed for accurate voxel placement.
440 Metabolite spectra were acquired with water suppression (water saturation pulse with
441 bandwidth of 50 Hz) and 128 signal averages. Additionally, 6 water reference scans were
442 acquired. Total acquisition time, including prescans, was approximately 5 min, and all raw data
443 were archived for processing offline.

444 Pre- and post-intervention BOLD resting-state scan

445 Participants were instructed to rest, breathe normally and look at the central white cross on
446 the screen.

447 Pre- and post-intervention pCASL resting-state scan

448 To assess whether the intervention affected blood flow during rest, in both MR sessions
449 participants completed a second short resting-state scan. Participants were instructed to rest
450 while breathing normally with their eyes open. To make visual inputs like those viewed during
451 the training scan (for our analyses comparing rest vs. training scans), we presented red and
452 blue circles alternately at a random rate (see Training sessions during BOLD and pCASL section
453 below). Participants were asked not to pay attention to these stimuli.

454 Training mimicking session during BOLD and pCASL

455 In the post-intervention scan session after the resting-state and emotion-regulation scans,
456 participants completed their daily training without biofeedback during BOLD and pCASL scans.
457 By this point, participants were well-trained, having each completed on average 57 training
458 sessions at home. For the Osc+ group, a red and blue circle alternated at their resonance
459 frequency. For example, if their resonance frequency was 12 sec, the red circle was presented
460 for 6 sec followed by the blue circle for 6 sec. Participants were asked to breathe in with the
461 red circle and breathe out with the blue circle. For the Osc- group, the stimuli were the same
462 as for the Osc+ group; however, the red and blue circles alternated at a random rate and
463 participants were told not to pay attention to them. See Fig. 3 for a visual representation of
464 training mimicking session.

465 Emotion regulation task

466 Participants completed an emotion regulation task³⁷ in the MRI scanner, which lasted for
467 about 10 min. Each trial consisted of three parts: instruction (1s), regulation (6s), and rating
468 (4s). First, participants were given one of three instructions: “view”, “intensify,” or “diminish.”
469 Then, during the regulation phase, they saw a positive, neutral, or negative image. Finally, they
470 were asked to rate the strength of the feeling they were experiencing on a scale ranging from

471 1 (very weak) to 4 (very strong). See Fig. 3 for a visual representation of emotion regulation
472 task.

473 Ultimatum Game task

474 During the week 7 MRI scanning session, participants completed an Ultimatum Game (UG)
475 task³⁸. Before scanning, participants were instructed that in this task they would be presented
476 by offers proposed by other participants of the study (with each player making 4 offers) or
477 offers randomly generated by the computer. Participants had the options to accept or reject
478 the offers. If they accepted an offer, the money would be split as proposed by the other player.
479 If they rejected an offer, both players would receive nothing on that trial. Participants were
480 told that at the end of the study, one of the trials in the game would be randomly selected and
481 both them and the proposer for that trial would be paid based on the participant's response.
482 In order to enhance realism, participants also played the role of proposer and were asked to
483 make offers to future participants. The task lasted for about 10 minutes and consisted of 36
484 trials. Out of these, 18 included fair and 18 included unfair offers. Fair offers ranged from 0.40
485 to 0.55 of the endowment whereas unfair offers were defined as ones ranging from 0.05 to
486 0.20 of the endowment. Each trial lasted 14 seconds. At the beginning of each trial, the face
487 and initials of the proposer for that trial were presented for a duration of 2s. Then the offer
488 proposed along with the face and initials of the proposer were shown for 5s. The decision
489 period was then followed, during which participants had 5s to respond to the offer.
490 Participants pressed one of two buttons on a button box to express their decision. Finally, the
491 results screen was presented for a jittered duration of 2-6s. In between trials, a fixation cross
492 was shown on screen which lasted between 1-4s. See Fig. 3 for a visual representation of
493 Ultimatum Game task.

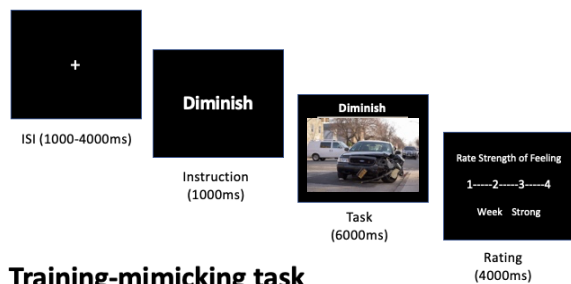
494 Younger and older adults completed two slightly different versions of the task. In younger
495 adults' version of the task, half of the human proposers were proposers that players knew,
496 while the other half were strangers. For familiar proposers, pictures of participants' group
497 mates were used. In older adults' version of the task, all human proposers were strangers.

498 Additional physiological measures during MRI

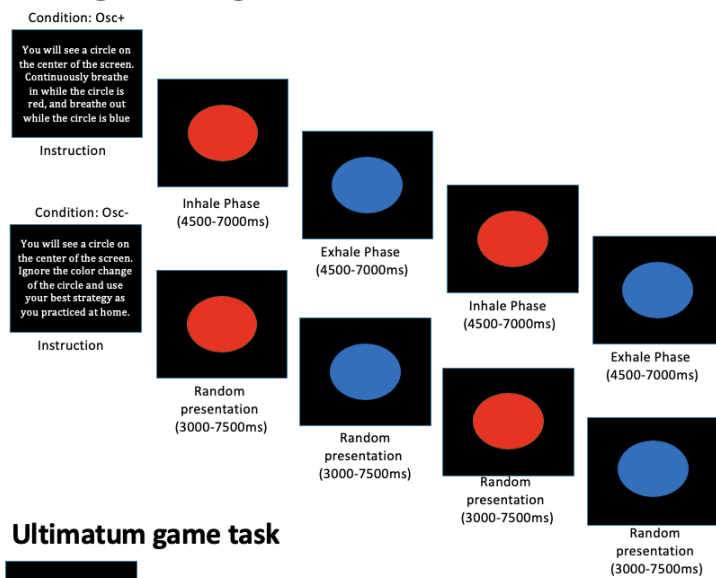
499 The physiological data collected during MRI scans include respiration, exhaled carbon dioxide
500 (CO₂), electrodermal activity, blood pressure, and heart rate. All the physiological data were
501 collected at 10kHz sampling rate using Biopac MP150 Data Acquisition System with MR-
502 compatible sensors and recorded with AcqKnowledge software 5.0. Respiration was measured
503 using the breathing belt, TSD201 transducer and transferred to the Biopac RSP100C module
504 to be 0.05-1Hz bandpass-filtered, amplified with 10 times of gain. Exhaled carbon dioxide
505 (CO₂) levels were measured using Philips NM3 Monitor (Model 7900) with nasal cannula and
506 fed to Biopac MP150. The heart rate data were collected with a Nonin Medical 8600FO Pulse
507 Oximeter and sent to Biopac MP150. The electrodermal activity was recorded using the Biopac
508 GSR100C module. Blood pressure was measured using CareTaker device and software and
509 recorded with Biopac MP150.

510

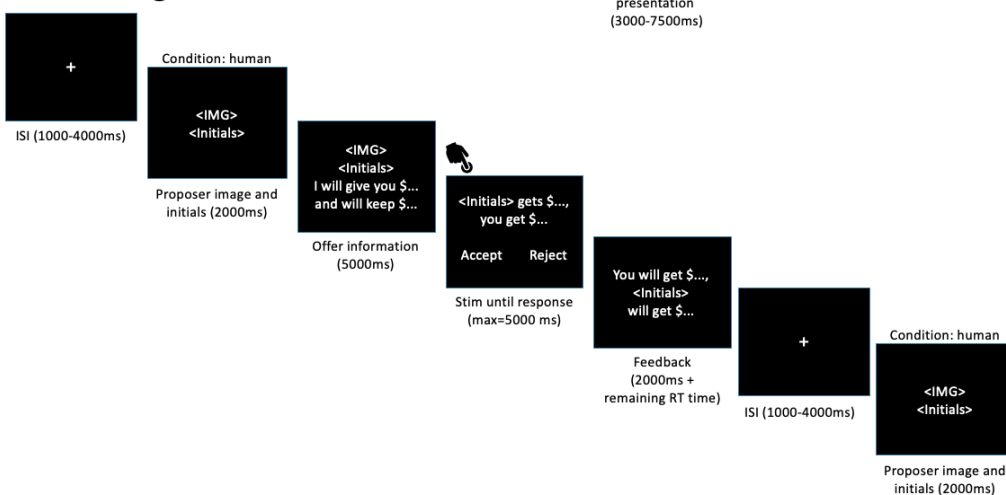
Emotion regulation task



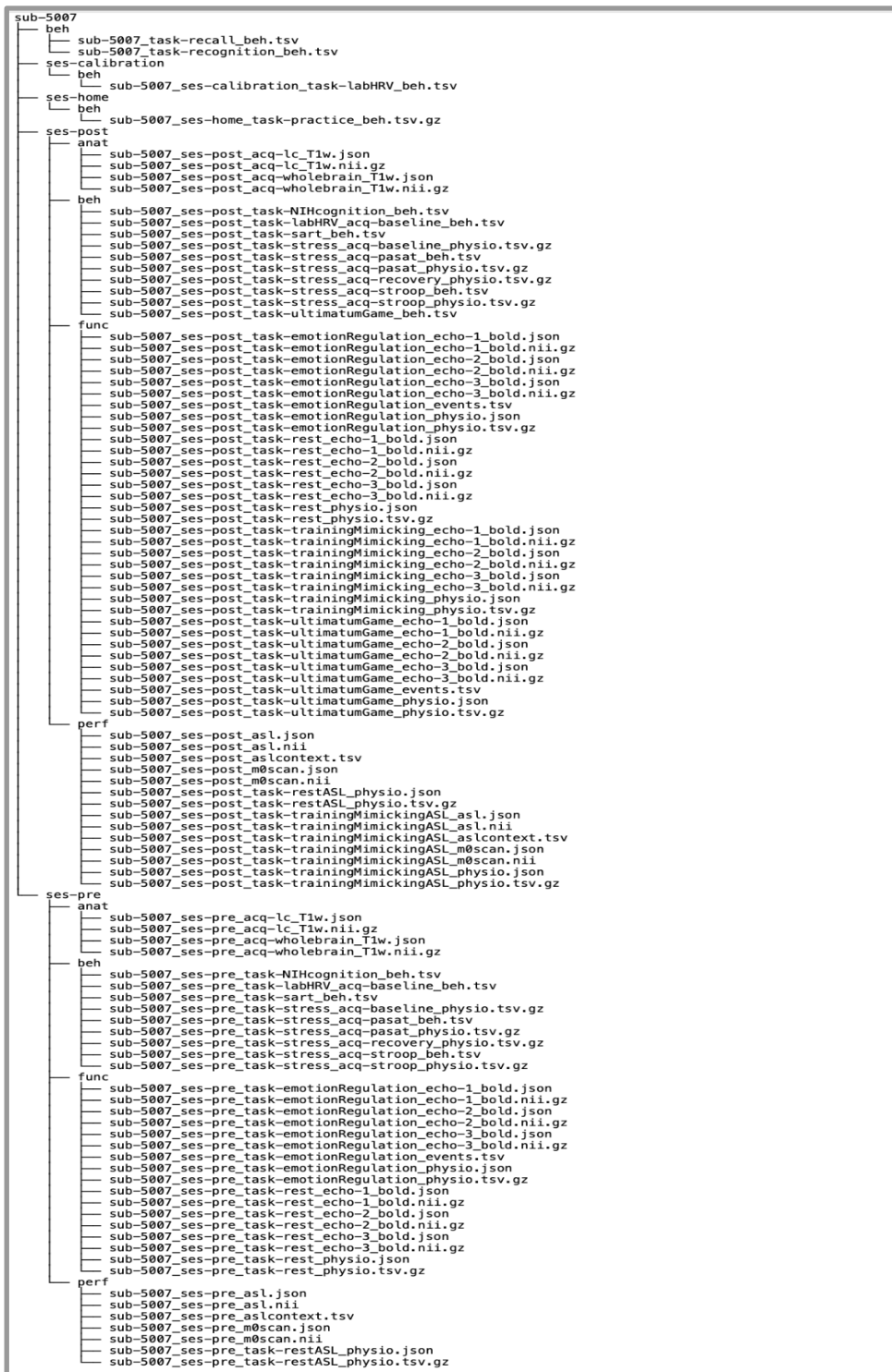
Training-mimicking task



Ultimatum game task



511 Fig. 3. A visual representation of all experimental paradigms during task-based fMRI. ISI:
512 inter-stimulus interval.



513 **Fig. 4. Example of the BIDS data structure for one participant.** The data for subject 5007 are
 514 organized into five folders; two sessions for pre and post measurements, two sessions for HRV
 515 biofeedback data, one for calibration and the other for home training, and one last folder for
 516 behavioural data. While the data structure is consistent across subjects, there is some
 517 variation regarding data availability.

518 **Data Records**

519 The following data are available on the OpenNeuro data sharing platform
520 (<https://openneuro.org/datasets/ds003823>). The files were organized in Brain Imaging Data
521 Structure (BIDS) format³⁹ (version 1.5.0; <http://bids.neuroimaging.io>). The BIDS provides a
522 convention of fMRI data naming and organization in order to facilitate the transfer, storage,
523 and sharing of neuroimaging data. The BIDS validation tool provided by OpenNeuro was used
524 to ensure that the dataset followed the BIDS system.

525 At the root level of the dataset, participant demographic information, including sex, and
526 handedness, and age group is provided in the participants.tsv file and these variables are
527 further described in the accompanying data dictionary, participants.json. The participants.tsv
528 file also indicates which of the different tasks, physiological data, and MRI scans are available
529 for each participant at each time point. This information is organized into 33 columns
530 containing “1” (data exist) or “0” (missing data) for all measures at each session (i.e., ses-
531 pre_task-emotionRegulation; ses-post_task-emotionRegulation). Also, we organized the
532 information about data quality in 15 columns containing “1” (recommend excluding) or “0”
533 (recommend including) for MRI or physiological measures at each session.

534 We organized the rest of the participants’ data in three ways: phenotype, subject folders, and
535 derivatives folder.

536 1) “Phenotype”: This folder includes files that list all participants’ scores on standardized tests
537 at each time point and participants’ responses to emotion questionnaires at each time-point
538 (with one row per participant)

539 2) “Sub- < ID >”: This folder contains participants’ MRI scan data, physiological measures, and
540 behavioral measures. Inside the folder of each participant with data available (i.e., n = 193 for
541 total participants and n = 162 for longitudinal data, see also Fig. 1 for detailed information),
542 there are two subfolders, named “ses-pre”, “ses-post” that contain data collected during pre-
543 and post-intervention sessions, respectively. Another two subfolders, named “ses-calibration”,
544 “ses-home,” contain heart rate measures collected during in-lab calibration sessions and
545 during home practice sessions, respectively. The last subfolder, named “beh” has individual
546 data for the picture memory tasks, which were administered in Weeks 4 and 5.

547 Inside “ses-pre” and “ses-post” folders, there are four subfolders named “anat”, “func”, “perf”,
548 and “beh”. “Anat” folder contains T₁-weighted structural images, “func” folder contains multi-
549 echo BOLD scan data and the physiological data collected during those scans, “perf” folder
550 contains pCASL scan data and the physiological data collected during those scans, and “beh”
551 folder contains behavioral and physiological measures collected outside the scanner. Inside
552 the “func” subfolder, there are files containing participant’s performance on the task (i.e.,
553 ‘events’ file), and physiological data for each task (i.e., ‘physio’ file) in addition to brain image
554 files. The events file includes one row per trial, including onset time of each trial, duration of
555 the event, trial type, response, response time, and the presented stimulus. Inside the “perf”
556 subfolder, there are files containing (1) 10 tag & control acquisitions from the pCASL scan in a
557 4D file (“*_asl.nii”) and (2) an M0 calibration image from the pCASL scan in a 3D file
558 (“*_m0scan.nii”). Fig. 4 provides an example of the BIDS data structure for one subject. Table

559 2 provides detailed information about the file name and the location for each measure at each
560 time point.

561 3) “derivatives”: This folder contains MRS, mriqc, and freesurferQC folders. Inside the MRS
562 folder, there is a “MRS_summary.tsv” file and individual subject folders. “MRS_summary.tsv”
563 includes the individual metabolite concentration levels and quality metrics for all scans for all
564 participants. Inside each subject folder, there are two subfolders, named “ses-pre”, “ses-post”
565 including raw MRS data as IMA file format. The pre session included one MRS scan, which
566 produced three ‘.IMA’ files. During the post session, some participants had two MRS scans;
567 the first MRS scan occurred at the same point in the scan sequence as the pre MRS scan. The
568 second, optional, MRS scan was completed after all other task scans were done. Inside the
569 mriqc folder, there are multiple files for scan types; “group_T1w_mriqc.tsv” and
570 “group_BOLD_mriqc_<task-name>.tsv” include the quality control metrics for the T₁-weighted
571 and functional (BOLD) MRI scans, respectively. Inside the freesurferQC folder, there is a
572 freesurfer_QC.tsv file including Freesurfer quality metrics and outlier participants on these
573 metrics flagged.

574 **Table 2. Summary of data file name and location.**

Category	Data type	Data summary file		Description
		Individual file name for time-point1	Individual file name for time-point2	
Demographics	Demographics	•Participants.tsv		•Participants' identification number, sex, age, handedness, and data present at each task at each time-point
Questionnaire	self reported behavior and emotion	•phenotype/questionnaires_summary.tsv		•Summary of responses to emotion questionnaires (SAI, TAI, CESD, POMS, FFMQ, SRSI, CFQ, ERQ, PSS-10) and altruism scale
Sleep	Sleep time	•phenotype/Whoop_Summary.tsv		•Sleep and HRV derived from slow-wave sleep, measured with WHOOP wristbands.
Cognitive task	NIH toolbox-Cognition	•phenotype/NIH_Cognition_summary.tsv		•Performance on three tasks from NIH toolbox-Cognition; flanker test, list sorting working memory test, and pattern comparison processing speed test
		•sub-<ID>/ses-pre/beh/sub-<ID>_ses-pre_task-NIHcognition_beh.tsv	•sub-<ID>/ses-pre/beh/sub-<ID>_ses-post_task-NIHcognition_beh.tsv	
	SART	•phenotype/SART_summary.tsv		•Performance on Sustained Attention to Response Task
	Memory task	•sub-<ID>/beh/sub-<ID>_task-recall_beh.tsv •sub-<ID>/beh/sub-<ID>_task-recognition_beh.tsv		•Recall and recognition on picture memory tasks at week4 and week5 lab visit
HRV-biofeedback	calibration	•sub-<ID>/ses-calibration/beh/sub-<ID>_ses-calibration_task-labHRV_beh.tsv		•Heart rate data during lab calibration sessions; resting and several conditions included for each calibration session
	Home training	•sub-<ID>/ses-home/beh/sub-<ID>_ses-home_task-practice_beh.tsv		•Heart rate data during home practice
Stress task	Physiological data	•sub-<ID>/ses-pre/beh/sub-<ID>_ses-pre_task-stress_acq-baseline_physio.tsv.gz •Same for acq-pasat, stroop, and recovery	•sub-<ID>/ses-post/beh/sub-<ID>_ses-post_task-stress_acq-baseline_physio.tsv.gz •Same for acq-pasat, stroop, and recovery	•Physiological data (ECG, respiration, continuous blood pressure, and skin conductance response) to assess reactivity to and recovery from acute stress
	Behavioral data	•sub-<ID>/ses-pre/beh/sub-<ID>_ses-pre_task-stress_acq-pasat_beh.tsv •sub-<ID>/ses-pre/beh/sub-<ID>_ses-pre_task-stress_acq-stroop_beh.tsv	•sub-<ID>/ses-post/beh/sub-<ID>_ses-post_task-stress_acq-pasat_beh.tsv •sub-<ID>/ses-post/beh/sub-<ID>_ses-post_task-stress_acq-stroop_beh.tsv	• Behavioral responses during PASAT and stroop task to induce acute stress
MR Imaging	Functional-resting	•sub-<ID>/ses-pre/func/sub-<ID>_ses-pre_task-rest_echo-1_bold.nii.gz •Same for echo-2 and echo-3	•sub-<ID>/ses-post/func/sub-<ID>_ses-post_task-rest_echo-1_bold.nii.gz •Same for echo-2 and echo-3	•Multi-echo fMRI •BOLD resting-state scan
	Functional-resting-state ASL	•sub-<ID>/ses-pre/perf/sub-<ID>_ses-pre_asl.nii	•sub-<ID>/ses-post/perf/sub-<ID>_ses-post_asl.nii	• pCASL resting-state scan
	Functional-Emotion regulation	•sub-<ID>/ses-pre/func/sub-<ID>_ses-pre_task-emotionRegulation_echo-1_bold.nii.gz •Same for echo-2 and echo-3	•sub-<ID>/ses-post/func/sub-<ID>_ses-post_task-emotionRegulation_echo-1_bold.nii.gz •Same for echo-2 and echo-3	•Multi-echo fMRI data •Emotion regulation task during BOLD scan
	Anatomical-T ₁	•sub-<ID>/ses-pre/anat/sub-<ID>_ses-pre_acq-wholebrain_T1w.nii.gz	•sub-<ID>/ses-post/anat/sub-<ID>_ses-post_acq-wholebrain_T1w.nii.gz	•T ₁ -weighted structural scan
	Biochemistry-MRS	•derivatives/MRS/MRS_summary.tsv		•proton magnetic resonance spectroscopy
		•/derivatives/MRS/sub-<ID>/ses-pre/*.IMA	•/derivatives/MRS/sub-<ID>/ses-post/*.IMA	
	Anatomical-TSE	•sub-<ID>/ses-pre/anat/sub-<ID>_ses-pre_acq-lc_T1w.nii.gz	•sub-<ID>/ses-post/anat/sub-<ID>_ses-post_acq-lc_T1w.nii.gz	•two-dimensional, multi-slice TSE scan
	Functional-UG		•sub-<ID>/ses-post/func/sub-<ID>_ses-post_task-ultimatumGame_echo-1_bold.nii.gz •Same for echo-2 and echo-3	•Multi-echo fMRI data •Ultimatum Game task during BOLD scan
Functional-training mimicking		•sub-<ID>/ses-post/func/sub-<ID>_ses-post_task-trainingMimicking_echo-1_bold.nii.gz •Same for echo-2 and echo-3	•Multi-echo fMRI data •Training mimicking session during BOLD scan	
Functional-training mimicking ASL		•sub-<ID>/ses-post/perf/sub-<ID>_ses-post_task-trainingMimickingASL_m0scan.nii	•Training mimicking session during pCASL scan	

575

576 **Technical Validation**

577 In this section, we describe quality control metrics for each measure. We take a conservative
578 approach for data exclusions. We generally did not exclude brain imaging data unless (1) the
579 MRI scan session was interrupted by unexpected events (e.g., an earthquake or power outage),
580 (2) an absence of a usable T_1 -weighted scan due to technical error or scan terminated by
581 participants, or (3) incidental findings. Also, we did not exclude behavioral or physiological
582 data unless (1) the task was interrupted by unexpected events (e.g., an earthquake or power
583 outage) or (2) obvious sensor error or data input error due to a technical issue. But we applied
584 quality control for data analyses and shared the results of quality control metrics in the
585 derivative folder and quality control results in the participants.tsv file. This way, the future
586 users of the datasets can use the quality-controlled data we recommend including, evaluate
587 our quality control methods, or apply their own quality control methods on the datasets.
588 Importantly, this places the responsibility for inclusion and exclusion of data in the hands of
589 the users of the datasets.

590 **MRI Data quality assessment**

591 The quality control metrics for the T_1 -weighted and functional (BOLD) MRI scans were
592 computed by the MRIQC package, which outputs several quality control metrics as well as a
593 report with visualizations of different aspects of the data. The quality control metrics for T_1 -
594 weighted images are stored in the group_T1w.tsv in the derivatives/mriqc folder. The quality
595 control metrics for the functional scans are stored in derivatives/mriqc folder.

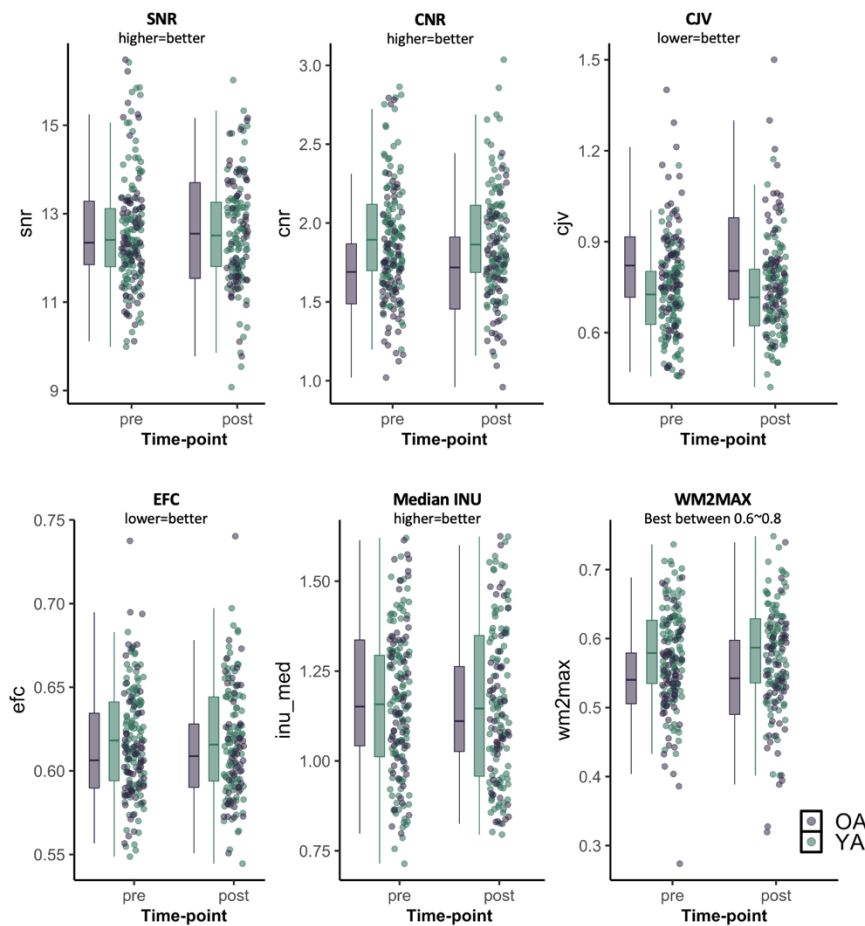
596 T_1 -weighted scans

597 All T_1 -weighted scans were run through the MRIQC pipeline, which outputs several quality
598 control metrics as well as a report with visualizations of different aspects of the data. All
599 individual subject reports were visually checked for artifacts including reconstruction errors,
600 failure of defacing, and segmentation. Defacing was considered to be successful if the 3D
601 render did not contain more than one partial facial feature (eyes, nose, or mouth) and no brain
602 tissue had been removed during defacing⁴⁰. In Fig. 5, we visualize several quality control
603 metrics related to the T_1 -weighted scans over two time-points (pre and post). MRIQC includes
604 the signal-to-noise ratio (SNR) calculation proposed by Dietrich et al.⁴¹, using the air
605 background as a noise reference. Additionally, for images that have undergone some noise
606 reduction processing, or the more complex noise realizations of current parallel acquisitions,
607 a simplified calculation using the within tissue variance is also provided. Higher values indicate
608 better quality. The contrast-to-noise ratio (CNR)⁴² is an extension of the SNR calculation to
609 evaluate how separated the tissue distributions of GM and WM are. Higher values indicate
610 better quality. The coefficient of joint variation (CJV) of GM and WM was proposed as objective
611 function⁴³ for the optimization of intensity non-uniformity (INU) correction algorithms. Higher
612 values are related to the presence of heavy head motion and large INU artifacts. The entropy-
613 focus criterion (EFC)⁴⁴ uses the Shannon entropy of voxel intensities as an indication of
614 ghosting and blurring induced by head motion. Lower values are better. Median INU is an
615 index of spatial inhomogeneity. It estimates the location and spread of the bias field
616 extracted⁴⁵. The smaller spreads located around 1.0 are better. The white matter to maximum
617 intensity ratio (WM2MAX) is the median intensity within the WM mask over the 95%

618 percentile of the full intensity distribution, that captures the existence of long tails due to
619 hyper-intensity of the carotid vessels and fat. Values should be around the interval [0.6, 0.8]⁴⁶.
620 In general, data quality appears consistent across time. All quality control metrics related to
621 the T₁-weighted scans for each participant, including those visualized in Fig. 5, are stored in
622 the group_T1w_mriqc.tsv file in the derivatives/mriqc folder. Here we do not exclude any
623 subjects based on IQMs, but subsequent researchers can use the available IQMs to exclude
624 scans as they see fit.

625 Each participant's T₁-weighted structural images were preprocessed using Freesurfer image
626 analysis suite version 6.0 (<http://surfer.nmr.mgh.harvard.edu/>). Cortical reconstruction and
627 volumetric segmentation were performed. Following initial preprocessing, we used the
628 Freesurfer 6.0 image analysis suite longitudinal stream to automatically extract volume
629 estimates⁴⁷. After completing the longitudinal Freesurfer pipeline, we used automated
630 measures computed by FreeSurfer of the contrast-to-noise ratio (the difference in signal
631 intensity between regions of different tissue types and noise signal) and the Euler number (a
632 metric of cortical surface reconstruction) to identify poor quality structural scans (Chalavi et
633 al. 2012; Rosen et al. 2018). For analyses of volumetric change, we identified outliers (N = 4
634 for younger adults and N = 2 for older adults) who on a box-and-whisker plot were above Q3
635 + 3 × the interquartile range on either of these metrics on either pre or post images. Freesurfer
636 quality metrics and the list of outliers are provided in the freesurfer_QC.tsv file in the
637 derivatives/freesurferQC folder.

638

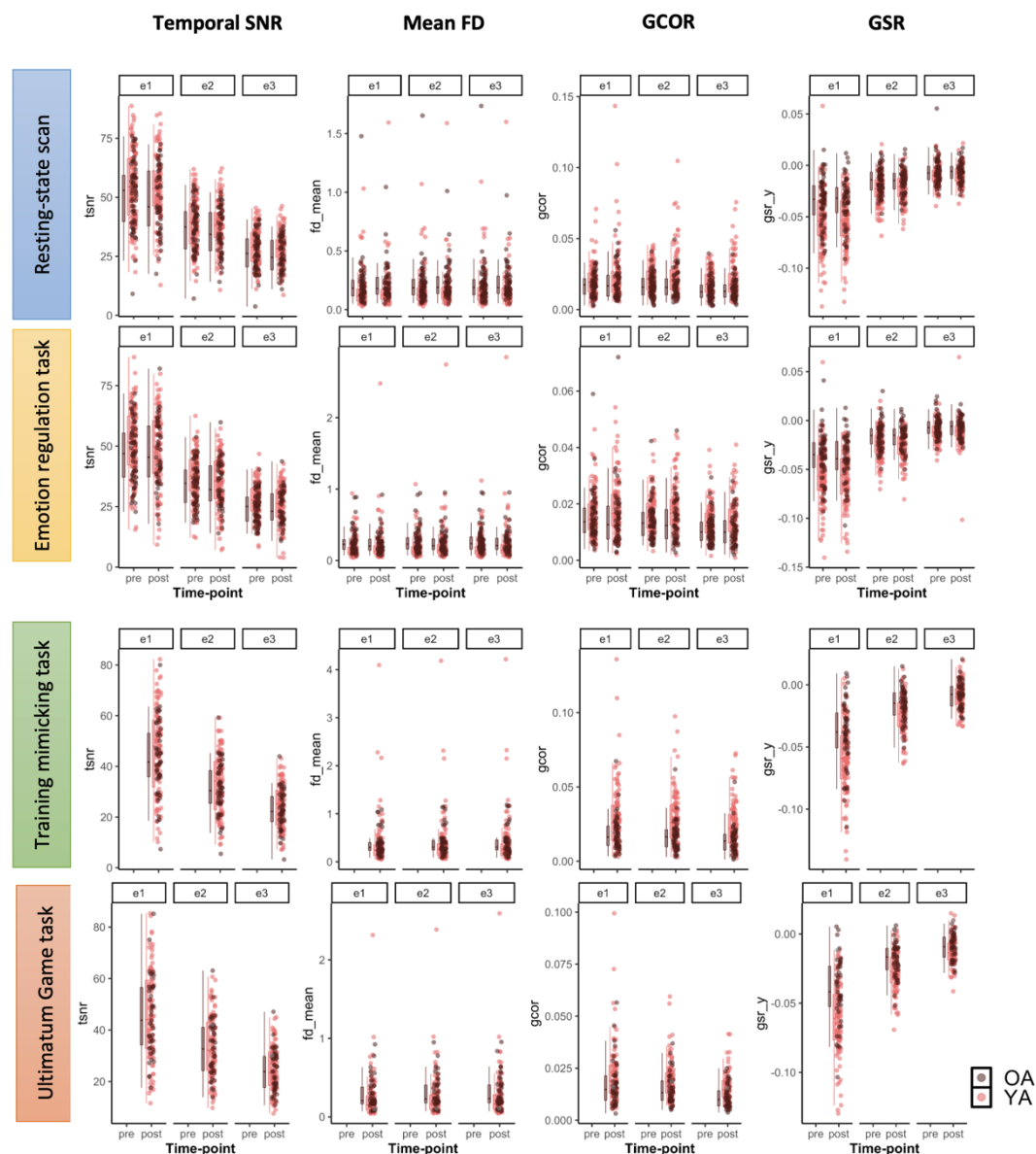


639 **Fig. 5. Quality control metrics related to the T₁-weighted scans at each time-point. Quality**
 640 **control metrics related to the T₁-weighted scans.** SNR: signal-to-noise ratio; CNR: contrast-to-
 641 noise ratio; CJV: coefficient of joint variation, an index reflecting head motion and spatial
 642 inhomogeneity; EFC: entropy-focused criterion, an index reflecting head motion and ghosting;
 643 Median INU: intensity non-uniformity, an index of spatial inhomogeneity; WM2MAX: ratio of
 644 median white-matter intensity to the 95% percentile of all signal intensities

645 Functional (BOLD) scans

646 The functional (BOLD) scans were run through the MRIQC pipeline. The resulting reports were
 647 visually checked for artifacts including reconstruction errors, registration issues, and incorrect
 648 brain masks. In Fig. 6, we visualize several quality control metrics related to the functional
 649 scans across three echo times (e1 = 18ms, e2 = 35, and e3 = 53 ms) over two time-points (pre
 650 and post). Temporal SNR (tSNR) is a simplified interpretation of the tSNR definition⁴⁸. The
 651 MRIQC pipeline provided the median value of the tSNR map calculated as, $tSNR = \langle S \rangle_t / \sigma_t$,
 652 where $\langle S \rangle_t$ is the average BOLD signal (across time), and σ_t is the corresponding temporal
 653 standard-deviation map. Higher values are better when comparing scans at the same echo
 654 (differences across echo times are expected due to effects of echo time on BOLD contrast).
 655 Mean Framewise Displacement (FD) is a measure of subject head motion, which compares the
 656 motion between the current and previous volumes. Higher values indicate lower quality.

657 Global Correlation (GCOR) is the average correlation of all pairs of voxel time series inside of
 658 the brain. GCOR measures differences between data due to motion/physiological
 659 noise/imaging artifacts as well as global neural fluctuations^{49,50}. MRIQC measures ghost-to-
 660 signal ratio (GSR) along the x or y encoding axes. Higher values indicate lower quality. Like the
 661 T₁-weighted quality control metrics, the functional quality metrics appear consistent across
 662 time. All quality control metrics related to the functional (BOLD) scans for each participant,
 663 including those visualized in Fig. 6, are provided in the group_BOLD_mriqc.tsv file in the
 664 derivatives/mriqc folder. Here we do not exclude any subjects based on IQMs, but subsequent
 665 researchers can use the available IQMs to exclude scans as they see fit.



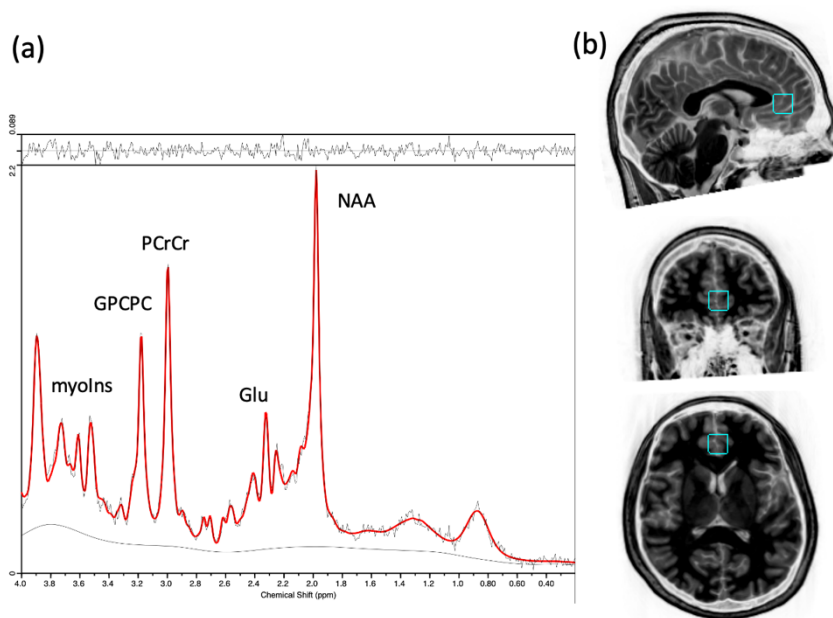
666 **Fig. 6. Quality control metrics related to the multi-echo functional (BOLD) scans at each time-**
 667 **point for resting-state scan, emotion regulation task scan, training mimicking task scan, and**
 668 **UG task.** SNR: signal-to-noise ratio, an index of signal quality; FD: framewise displacement, an
 669 index of overall movement; GCOR: global correlation, an index of the presence of global signals;
 670 GSR: ghost-to-signal ratio, an index of ghosting along the phase-encoding axis.

671 MRS scans

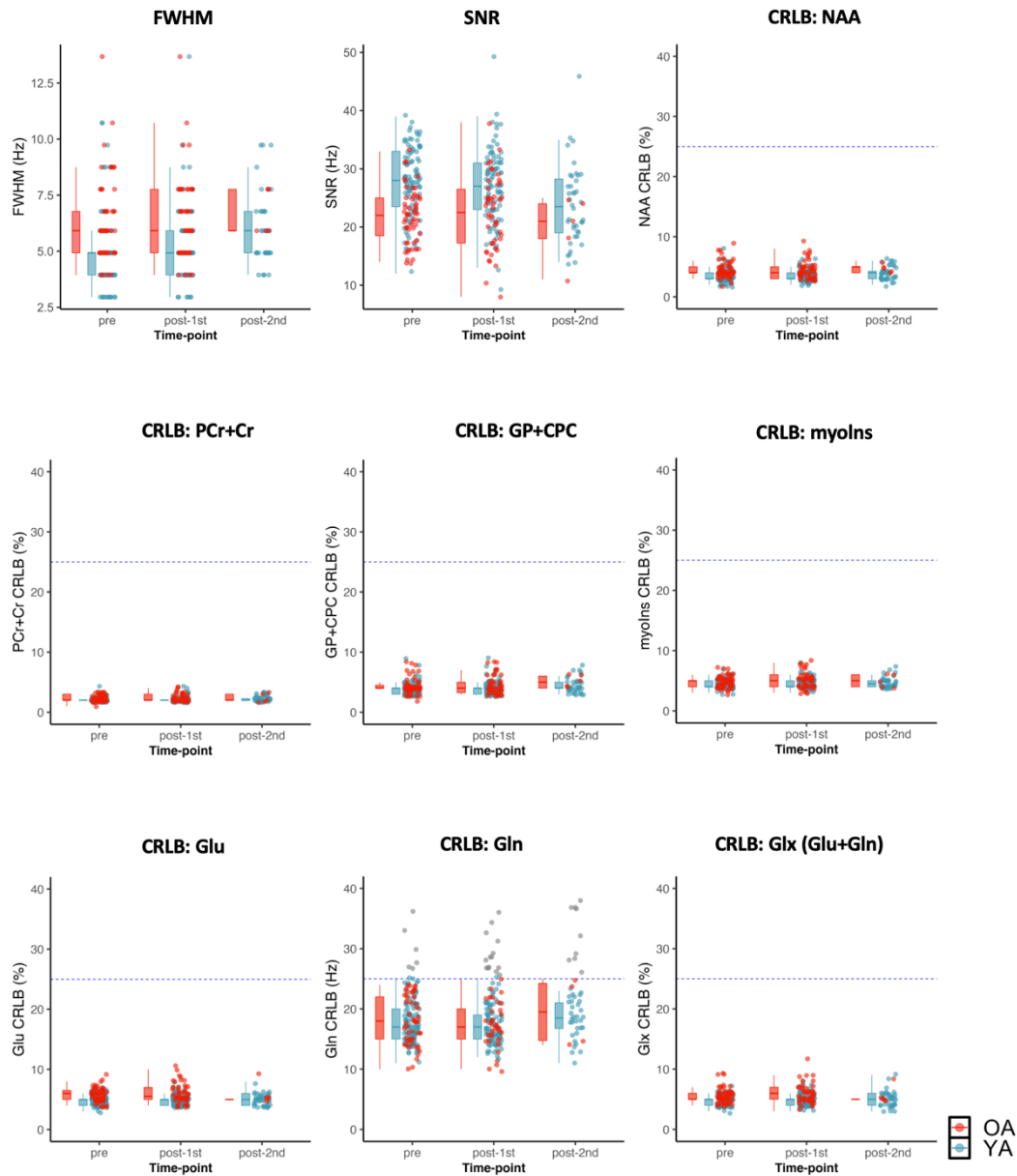
672 Post-processing and quantification of the ^1H MRS data were 100% automated⁵¹. For each ^1H
673 MRS spectra, the metabolites N-acetyl-aspartate (NAA), phosphocreatine plus creatine
674 (PCr+Cr), trimethylamines [glycerophosphocholine plus phosphocholine (GP+CPC)], and myo-
675 inositol, glutamate, and glutamine (as well as the less reliable metabolites, aspartate, gamma-
676 aminobutyric acid, glutathione, lactate, n-acetylaspartylglutamate, scyllo-inositol and taurine)
677 were quantified using the Linear Combination (LC) Model software⁵² with a simulated basis set
678 for the a priori knowledge reflecting the acquisition parameters. An example of an individual
679 MRS spectrum from the ^1H MRS voxel placed in the anterior cingulate cortex is shown in Fig.
680 7. Freesurfer and FSL tools (FLIRT, FAST, MRI_VOLSYNTH, MRI_VOL2VOL) were used to tissue
681 segment the T_1 -weighted images, which were then used to quantify the tissue fraction values
682 within each voxel location.

683

684 As quality control metrics for the MRS scans, we used SNR, line width reflecting the full width
685 at half maximum (FWHM) of NAA, and Cramér-Rao lower bound (CRLB)⁵³. Nine MRS spectra
686 were rejected for poor quality (pre: 3, post-1st: 3, post-2nd: 3) due to extreme CRLB values
687 and the distribution of quality metrics are visualized in Fig. 8 after removal of poor quality data.
688 Also, all metabolite levels that have CRLB higher than 25% or another chosen threshold are
689 tagged “outlier” with gray color in Fig. 8. But we included all data with quality tag in the shared
690 data file, “MRS_summary.tsv”, for future users of the datasets to apply their own threshold
691 on the datasets⁵⁴. The individual metabolite levels and quality metrics are provided for all
692 scans for all participants in the MRS_summary.tsv file in the derivatives/MRS folder.



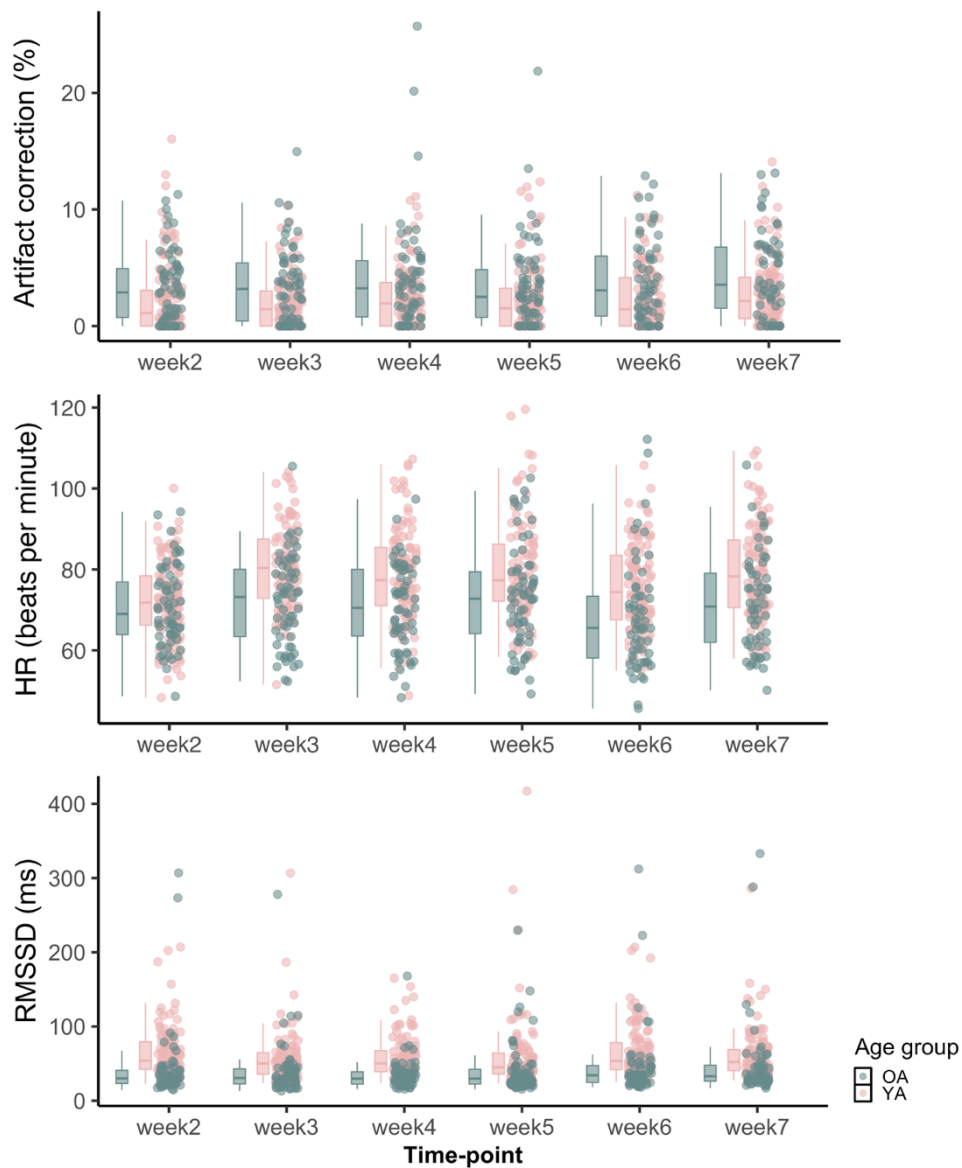
693 **Fig. 7. An example of an individual quantified MRS spectrum (a) and the sagittal, coronal and**
694 **axial view of the MRS voxel placed in the anterior cingulate cortex, from top to bottom (b).**



695 **Fig. 8. Quality control metrics related to the MRS scans at each time-point.** 9 out of 353
 696 spectra were rejected as bad quality and not included in the figure. CRLB >25% are detected
 697 as outliers specified by gray color. FWHM: full width at half maximum of singlet peaks; SNR:
 698 signal to noise ratio; CRLB: Cramér-Rao lower bound; NAA:N-acetyl-aspartate; PCr+Cr:
 699 phosphocreatine plus creatine ; GP+CPC: glycerophosphocholine plus phosphocholine;
 700 myoIns: myo-inositol; Glu: glutamate; Gln: glutamine; Glu_Gln: Glu + Gln.

701 **Resting heart rate data quality assessment**

702 The resting heart rate data was measured during weekly HRV calibration sessions. The pulse
703 was measured with an infrared pulse plethysmograph (ppg) ear sensor and the interbeat
704 interval (IBI) data was extracted after eliminating ectopic beats or other sources of artifacts
705 through a built-in process in emWave pro software. Fig. 9 depicts distributions of the artifact
706 correction rate, mean heart rate, and RMSSD.



707 **Fig. 9. Distribution of artifact correction (%), heart rate (HR), and RMSSD during rest from**
708 **weekly lab calibration session.**

709 **Behavioral measure quality assessment**

710 We calculated descriptive statistics and reliability estimates of each subscale of the emotion
711 questionnaire (Table 3). We reported average Cronbach’s alpha coefficients from multiple
712 time-points to ensure their internal consistency. To check the test-retest reliability, we
713 reported the intraclass correlation coefficients and typical percentage error⁵⁵.

714 **Table 3. Summary of internal consistency (Cronbach’s alpha) and test-retest reliability**
715 **estimates in emotion questionnaire.**

	Abbr.	Full Name	Reference	No. of Items	Construct	M (SD)	Cronbach’s alpha	Intraclass correlation coefficient	Typical percentage error (%)
1	POMS	The profile of mood states	Grove & Prapavessis (1992)	40	Total Mood Disturbance	88.46 (14.4)	0.88 (0.85 - 0.91)	0.67	11.02
2	SAI	The state anxiety inventory	Spielberger et al.(1983)	20	State anxiety	36.33 (9.36)	0.72 (0.59 - 0.77)	0.70	15.92
3	TAI	The trait anxiety inventory	Spielberger et al.(1983)	20	Trait anxiety	38.46 (10.04)	0.55 (0.46 - 0.61)	0.71	14.86
4	CESD	The Center for Epidemiological Studies Depression Scale	Radloff (1977)	20	Depression	12.44 (7.74)	0.71 (0.67 - 0.74)	0.71	37.33
5	FFMQ	The Five Facet Mindfulness Questionnaire	Tran et al. (2013)	20	Mindfulness	67.98 (10.33)	0.65 (0.60 - 0.70)	0.82	6.64
6	SRS13	The Smith Relaxation States Inventory 3-I	Smith (2001)	38 (26)	Relaxation (state)	3.11 (0.84)	0.94 (0.93 - 0.95)	0.71	15.41
				38 (8)	Stress (state)	1.81 (0.64)	0.83 (0.82- 0.83)	0.56	26.16
		The Smith Relaxation States Inventory 3-II		38 (26)	Relaxation (frequency)	3.78 (0.85)	0.93 (0.927- 0.932)	0.66	13.83
				38 (8)	Stress (frequency)	3.1 (0.96)	0.82 (0.818- 0.826)	0.72	17.33
7	CFQ 11	The Chalder Fatigue Scale	Jackson (2015)	11	Fatigue	12.48 (4.46)	0.86 (0.85 - 0.87)	0.56	25.53
8	ERQ	The Emotion Regulation Questionnaire	Gross & John (2003)	10	Emotion regulation (frequency)	45.40 (7.24)	0.75 (0.74 - 0.76)	0.67	10.15
				10	Emotion regulation (self-efficacy)	49.92 (8.96)	0.88 (0.86 - 0.90)	0.55	13.52
9	PSS-10	Perceived stress scale	Cohen et al., (1983)	10	Perceived stress	2.65 (0.59)	0.86 (0.84 - 0.89)	0.61	14.61

716

717 **Physiological measure quality assessment**

718 Table 4 summarizes the main characteristics of physiological data collected during 7-minute
719 resting-state BOLD MRI scans over two time-points. The original data files during MRI scans
720 were stored using the Biopac AcqKnowledge software and later converted as a tsv file for
721 sharing.

722 **Table 4. Descriptive statistics and test-retest correlation of physiological measures during**
723 **resting scan.**

Age group	Condition		Respiration rate (Hz)		Heart rate (bpm)		ETCO ₂ (mmHg)		
			week 2	week 7	week 2	week 7	week 2	week 7	
YA	Osc+	M	0.27	0.25	68.90	66.14	41.00	40.44	
		SD	0.06	0.07	9.65	8.69	3.13	3.27	
		N	54	46	59	45	35	41	
	Osc-	M	0.29	0.29	69.79	67.97	40.56	40.76	
		SD	0.06	0.06	10.33	9.65	4.02	3.99	
		N	43	41	51	43	31	37	
	Week 2 & 7 correlation			$r(68) = 0.69, p < 0.001$		$r(80) = 0.53, p < 0.001$		$r(44) = 0.50, p < 0.001$	
	OA	Osc+	M	0.23	0.22	65.87	62.11	39.97	40.11
			SD	0.07	0.06	9.55	10.95	4.78	4.42
N			28	25	29	25	28	24	
Osc-		M	0.22	0.23	70.35	69.34	41.07	41.65	
		SD	0.06	0.06	11.78	13.59	4.04	4.30	
		N	24	24	27	21	25	25	
Week 2 & 7 correlation			$r(40) = 0.81, p < 0.001$		$r(41) = 0.89, p < 0.001$		$r(42) = 0.80, p < 0.001$		

724

725 **Code Availability**

726 All code for collecting, formatting, and processing the data is available at
727 https://github.com/EmotionCognitionLab/HRV-ER-dataset_release. Information about the
728 code dependencies and package requirements are available in the same Github repository.

729

730 **Acknowledgements**

731 This study was supported by NIH R01AG057184 (PI Mather). We thank study participants for
732 their contribution. We thank the research assistant team for their help with data collection:
733 Michelle Wong, Kathryn Cassutt, Collin Amano, Heekyung Rachael Kim, Seungyeon Lee,
734 Alexandra Haydinger, Lauren Thompson, Gabriel Shih, Divya Suri, Sophia Ling, Akanksha Jain,
735 Linette Bagtas, Sumedha Attanti, Ivy Hsu, Michael Kwan, and Juliana Lee.

736

737 **Author contributions**

738 The authors made the following contributions. HY: Conceptualization, Data curation, Formal
739 analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original
740 draft; KN: Conceptualization, Data curation, Investigation, Writing - review & editing, Project
741 administration; JM: Conceptualization, Data curation, Investigation; CCho: Conceptualization,
742 Data curation, Investigation, Resources, Project administration; NM: Data curation, Software,
743 Validation; SB, PN, & SP: Conceptualization, Data curation, Investigation, Writing - review &
744 editing; SD: Investigation, Writing - review & editing; PC, YZ, & VG: Data curation, Investigation;
745 TF: Data curation, Software; JFT, PL, CChang, VM, SN, & DN: Conceptualization, Writing -
746 review & editing; JS: Data curation, Formal analysis, & Writing; JW: Conceptualization & Data
747 curation; MM: Conceptualization, Funding acquisition, Resources, Project administration,
748 Supervision, Writing - review & editing.

749

750 **Competing interests**

751 The authors have no relevant financial or non-financial interests to disclose.

752 References

753

- 754 1 Beauchaine, T. P. & Thayer, J. F. Heart rate variability as a transdiagnostic biomarker
755 of psychopathology. *Int J Psychophysiol* **98**, 338-350,
756 doi:10.1016/j.ijpsycho.2015.08.004 (2015).
- 757 2 Ge, F., Yuan, M., Li, Y. & Zhang, W. Posttraumatic stress disorder and alterations in
758 resting heart rate variability: A systematic review and meta-analysis. *Psychiatry*
759 *investigation* **17**, 9, doi:10.30773/pi.2019.0112 (2020).
- 760 3 Kemp, A. H. *et al.* Impact of depression and antidepressant treatment on heart rate
761 variability: a review and meta-analysis. *Biological psychiatry* **67**, 1067-1074,
762 doi:10.1016/j.biopsych.2009.12.012 (2010).
- 763 4 Sclocco, R. *et al.* The influence of respiration on brainstem and cardiovagal response
764 to auricular vagus nerve stimulation: a multimodal ultrahigh-field (7T) fMRI study.
765 *Brain stimulation* **12**, 911-921, doi:10.1016/j.brs.2019.02.003 (2019).
- 766 5 Goessel, V., Curtiss, J. & Hofmann, S. The effect of heart rate variability biofeedback
767 training on stress and anxiety: a meta-analysis. *Psychological Medicine* **47**, 1-9,
768 doi:10.1017/S0033291717001003 (2017).
- 769 6 Lehrer, P. *et al.* Heart Rate Variability Biofeedback Improves Emotional and Physical
770 Health and Performance: A Systematic Review and Meta Analysis. *Appl*
771 *Psychophysiol Biofeedback* **45**, 109-129, doi:10.1007/s10484-020-09466-z (2020).
- 772 7 Mather, M. & Thayer, J. How heart rate variability affects emotion regulation brain
773 networks. *Curr Opin Behav Sci* **19**, 98-104, doi:10.1016/j.cobeha.2017.12.017 (2018).
- 774 8 HeartMath®Institute. emWavePro. (2020).
- 775 9 Lehrer, P. *et al.* Protocol for heart rate variability biofeedback training. *Biofeedback*
776 **41**, 98-109 (2013).
- 777 10 Tarvainen, M. P., Niskanen, J. P., Lipponen, J. A., Ranta-Aho, P. O. & Karjalainen, P. A.
778 Kubios HRV--heart rate variability analysis software. *Comput Methods Programs*
779 *Biomed* **113**, 210-220, doi:10.1016/j.cmpb.2013.07.024 (2014).
- 780 11 Feng, T. Biofeedback to decrease HRV. (2018).
- 781 12 Grove, J. R. & Prapavessis, H. Preliminary Evidence for the Reliability and Validity of
782 an Abbreviated Profile of Mood States. *International Journal of Sport Psychology* **23**,
783 93-109 (1992).
- 784 13 Spielberger, C. D. State-trait anxiety inventory for adults. (1983).
- 785 14 Radloff, L. S. The CES-D scale: A self-report depression scale for research in the
786 general population. *Applied psychological measurement* **1**, 385-401 (1977).
- 787 15 Tran, U. S., Glück, T. M. & Nader, I. W. Investigating the Five Facet Mindfulness
788 Questionnaire (FFMQ): Construction of a short form and evidence of a two-factor
789 higher order structure of mindfulness. *Journal of Clinical Psychology* **69**, 951-965,
790 doi:10.1002/jclp.21996 (2013).
- 791 16 Smith, J. C. *Advances in ABC relaxation: Applications and inventories*. (Springer
792 Publishing Company, 2001).
- 793 17 Smith, J. C. RMM Tracker/SRSI3 MANUAL v9. 15.2020 Jonathan C. Smith, PhD.
- 794 18 Jackson, C. The Chalder fatigue scale (CFQ 11). *Occupational medicine* **65**, 86-86,
795 doi:10.1093/occmed/kqu168 (2015).
- 796 19 Gross, J. J. & John, O. P. Individual differences in two emotion regulation processes:
797 implications for affect, relationships, and well-being. *Journal of personality and social*
798 *psychology* **85**, 348, doi:10.1037/0022-3514.85.2.348 (2003).
- 799 20 Kupst, M. J. *et al.* Assessment of stress and self-efficacy for the NIH Toolbox for
800 Neurological and Behavioral Function. *Anxiety, Stress, & Coping* **28**, 531-544,
801 doi:10.1080/10615806.2014.994204 (2015).

- 802 21 Cohen, S., Kamarck, T. & Mermelstein, R. A global measure of perceived stress.
803 *Journal of health and social behavior* **24**, 385-396 (1983).
- 804 22 Rushton, J. P., Chrisjohn, R. D. & Fekken, G. C. The altruistic personality and the self-
805 report altruism scale. *Personality and individual differences* **2**, 293-302 (1981).
- 806 23 Dickerson, S. S. & Kemeny, M. E. Acute stressors and cortisol responses: a theoretical
807 integration and synthesis of laboratory research. *Psychological bulletin* **130**, 355,
808 doi:10.1037/0033-2909.130.3.355 (2004).
- 809 24 Tombaugh, T. N. A comprehensive review of the paced auditory serial addition test
810 (PASAT). *Archives of clinical neuropsychology* **21**, 53-76,
811 doi:10.1016/j.acn.2005.07.006 (2006).
- 812 25 MacLeod, C. M. Half a century of research on the Stroop effect: an integrative
813 review. *Psychological bulletin* **109**, 163, doi:10.1037/0033-2909.109.2.163 (1991).
- 814 26 Weintraub, S. *et al.* Cognition assessment using the NIH Toolbox. *Neurology* **80**, S54-
815 S64 (2013).
- 816 27 Gershon, R. C. *et al.* NIH toolbox for assessment of neurological and behavioral
817 function. *Neurology* **80**, S2-S6, doi:10.1212/WNL.0b013e3182872e5f (2013).
- 818 28 Rueda, M. R. *et al.* Development of attentional networks in childhood.
819 *Neuropsychologia* **42**, 1029-1040, doi:10.1016/j.neuropsychologia.2003.12.012
820 (2004).
- 821 29 Mungas, D., Reed, B. R., Farias, S. T. & Decarli, C. Criterion-referenced validity of a
822 neuropsychological test battery: equivalent performance in elderly Hispanics and
823 non-Hispanic Whites. *Journal of the International Neuropsychological Society* **11**,
824 620-630, doi:10.1017/S1355617705050745 (2005).
- 825 30 Crane, P. K. *et al.* Item response theory facilitated cocalibrating cognitive tests and
826 reduced bias in estimated rates of decline. *Journal of clinical epidemiology* **61**, 1018-
827 1027. e1019, doi:10.1016/j.jclinepi.2007.11.011 (2008).
- 828 31 Salthouse, T. A., Babcock, R. L. & Shaw, R. J. Effects of adult age on structural and
829 operational capacities in working memory. *Psychology and aging* **6**, 118,
830 doi:10.1037//0882-7974.6.1.118 (1991).
- 831 32 Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T. & Yiend, J. Oops!':
832 performance correlates of everyday attentional failures in traumatic brain injured
833 and normal subjects. *Neuropsychologia* **35**, 747-758 (1997).
- 834 33 Marchewka, A., Żurawski, Ł., Jednoróg, K. & Grabowska, A. The Nencki Affective
835 Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-
836 quality, realistic picture database. *Behavior research methods* **46**, 596-610,
837 doi:10.3758/s13428-013-0379-1 (2014).
- 838 34 Bellenger, C. R., Miller, D., Halson, S. L., Roach, G. & Sargent, C. Wrist-based
839 photoplethysmography assessment of heart rate and heart rate variability:
840 Validation of WHOOP. *Sensors* **21**, 3571, doi:10.3390/s21103571 (2021).
- 841 35 Miller, D. J. *et al.* A Validation Study of a Commercial Wearable Device to
842 Automatically Detect and Estimate Sleep. *Biosensors* **11**, 185,
843 doi:10.3390/bios11060185 (2021).
- 844 36 Chen, Y., Wang, D. J. & Detre, J. A. Test–retest reliability of arterial spin labeling with
845 common labeling strategies. *Journal of Magnetic Resonance Imaging* **33**, 940-949,
846 doi:10.1002/jmri.22345 (2011).
- 847 37 Kim, S. H. & Hamann, S. Neural correlates of positive and negative emotion
848 regulation. *Journal of cognitive neuroscience* **19**, 776-798,
849 doi:10.1162/jocn.2007.19.5.776 (2007).
- 850 38 Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E. & Cohen, J. D. The neural
851 basis of economic decision-making in the ultimatum game. *Science* **300**, 1755-1758,
852 doi:10.1126/science.1082976 (2003).

- 853 39 Gorgolewski, K. J. *et al.* The brain imaging data structure, a format for organizing and
854 describing outputs of neuroimaging experiments. *Sci Data* **3**, 160044,
855 doi:10.1038/sdata.2016.44 (2016).
- 856 40 Theyers, A. E. *et al.* Multisite Comparison of MRI Defacing Software Across Multiple
857 Cohorts. *Frontiers in psychiatry* **12**, 189, doi:10.3389/fpsy.2021.617997 (2021).
- 858 41 Dietrich, O., Raya, J. G., Reeder, S. B., Reiser, M. F. & Schoenberg, S. O. Measurement
859 of signal-to-noise ratios in MR images: influence of multichannel coils, parallel
860 imaging, and reconstruction filters. *Journal of Magnetic Resonance Imaging: An
861 Official Journal of the International Society for Magnetic Resonance in Medicine* **26**,
862 375-385, doi:10.1002/jmri.20969 (2007).
- 863 42 Magnotta, V. A. & Friedman, L. Measurement of signal-to-noise and contrast-to-
864 noise in the fBIRN multicenter imaging study. *Journal of digital imaging* **19**, 140-147,
865 doi:10.1007/s10278-006-0264-x (2006).
- 866 43 Ganzetti, M., Wenderoth, N. & Mantini, D. Intensity Inhomogeneity Correction of
867 Structural MR Images: A Data-Driven Approach to Define Input Algorithm
868 Parameters. *Frontiers in Neuroinformatics* **10**, 10, doi:10.3389/fninf.2016.00010
869 (2016).
- 870 44 Atkinson, D., Hill, D. L., Stoye, P. N., Summers, P. E. & Keevil, S. F. Automatic
871 correction of motion artifacts in magnetic resonance images using an entropy focus
872 criterion. *IEEE Transactions on Medical imaging* **16**, 903-910 (1997).
- 873 45 Esteban, O. *et al.* Crowdsourced MRI quality metrics and expert quality annotations
874 for training of humans and machines. *Scientific data* **6**, 1-7, doi:10.1038/s41597-019-
875 0035-4 (2019).
- 876 46 Esteban, O. *et al.* MRIQC: Advancing the automatic prediction of image quality in
877 MRI from unseen sites. *PloS one* **12**, e0184661, doi:10.1371/journal.pone.0184661
878 (2017).
- 879 47 Reuter, M., Schmansky, N. J., Rosas, H. D. & Fischl, B. Within-subject template
880 estimation for unbiased longitudinal image analysis. *Neuroimage* **61**, 1402-1418,
881 doi:10.1016/j.neuroimage.2012.02.084 (2012).
- 882 48 Krüger, G. & Glover, G. H. Physiological noise in oxygenation-sensitive magnetic
883 resonance imaging. *Magnetic Resonance in Medicine: An Official Journal of the
884 International Society for Magnetic Resonance in Medicine* **46**, 631-637,
885 doi:10.1002/mrm.1240 (2001).
- 886 49 Saad, Z. S. *et al.* Correcting brain-wide correlation differences in resting-state fMRI.
887 *Brain connectivity* **3**, 339-352 (2013).
- 888 50 Schölvinck, M. L., Maier, A., Ye, F. Q., Duyn, J. H. & Leopold, D. A. Neural basis of
889 global resting-state fMRI activity. *Proceedings of the National Academy of Sciences*
890 **107**, 10238-10243 (2010).
- 891 51 Stanley, J. A. *et al.* Functional dynamics of hippocampal glutamate during associative
892 learning assessed with in vivo 1H functional magnetic resonance spectroscopy.
893 *Neuroimage* **153**, 189-197 (2017).
- 894 52 Provencher, S. W. Estimation of metabolite concentrations from localized in vivo
895 proton NMR spectra. *Magnetic resonance in medicine* **30**, 672-679 (1993).
- 896 53 Sanaei Nezhad, F. *et al.* Quantification of GABA, glutamate and glutamine in a single
897 measurement at 3 T using GABA-edited MEGA-PRESS. *NMR in Biomedicine* **31**, e3847
898 (2018).
- 899 54 Kreis, R. The trouble with quality filtering based on relative Cramér-Rao lower
900 bounds. *Magnetic resonance in medicine* **75**, 15-18 (2016).
- 901 55 Hopkins, W. G. Measures of reliability in sports medicine and science. *Sports
902 medicine* **30**, 1-15, doi:10.2165/00007256-200030010-00001 (2000).
- 903