Supplementary Materials Papalini et al.

Stress matters: randomized controlled trial on the effect of probiotics on neurocognition

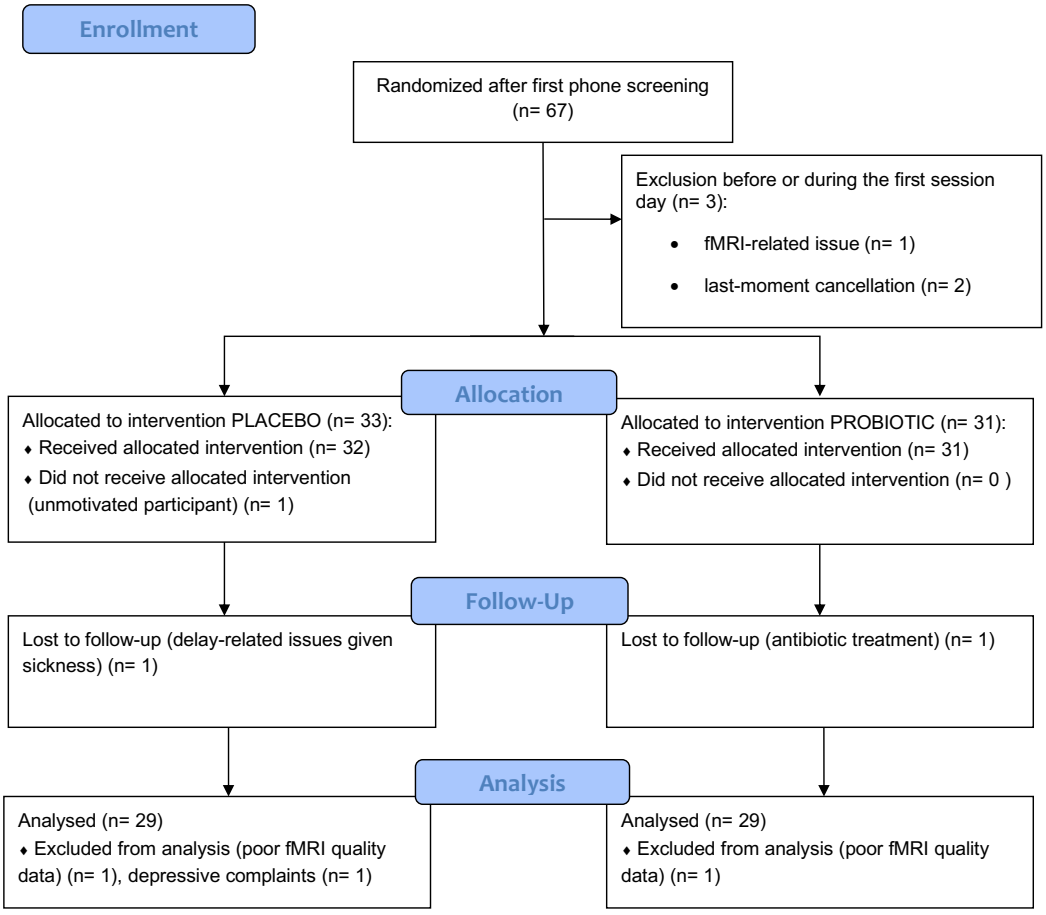
Participants

The aim of the study was to enroll sixty participants in total (30 for each placebo and probiotics group). To account for drop-outs, seven extra participants (for an initial total number of 67 participants) were additionally enrolled after a first phone screening. Three participants dropped out during the implementation of the study before the first testing session (i.e. last moment cancellations). After the first testing session, one participant was excluded given a lack of motivation in participating in the study, while two participants dropped out after/during the intervention before the second testing session (i.e. occurrence of sickness). Finally, one participant at risk of depression and two participants who obtained poor fMRI task performances (e. g. high miss and error rates), were excluded from the final analysis. This resulted in a final sample of 58 participants. See CONSORT flow diagram.

Diagram 1. *CONSORT flow diagram*

CONSORT
TRANSPARENT REPORTING of TRIALS

CONSORT 2010 Flow Diagram



Intervention details

Ecologic®barrier consisted of the following bacterial strains: *Bifidobacterium bifidum* W23, *Bifidobacterium lactis* W51, *Bifidobacterium lactis* W52, *Lactobacillus acidophilus* W37, *Lactobacillus brevis* W63, *Lactobacillus casei* W56, *Lactobacillus salivarius* W24, *Lactococcus lactis* W19 and, *Lactococcus lactis* W58. It has been confirmed that the probiotic formulation has always contained these nine strains and has not been changed in ratio or CFU count since it has been researched. The strains (total cell count of 2.5×10^9 colony forming units –cfu- per gram, i.e. 5×10^9 cfu per day) were blended into a carrier material consisting of maize starch, maltodextrin, vegetable protein and a mineral mix. The placebo consisted of the same carrier material as used in Ecologic®Barrier and was indistinguishable in color, smell, taste and appearance. With the application of new molecular identification techniques (including whole genome sequencing), the declaration of bacterial strains has been updated compared to previous publications (Steenbergen, Sellaro et al. 2015).

Questionnaires

BDI: The 21 items of the BDI are rated on a 4-point Likert scale (from 0 to 3 per item) indicating the severity of the feeling of the participant in the past two weeks. A total BDI score between 0-13 is an indicator of no-minimal depression. A BDI score above 13 (i.e. 14-19: mild depression; 20-28: moderate depression; 29-63: severe depression) was an exclusion criterion.

LEIDS-r: Before completing the LEIDS-r, participants were asked to imagine a situation when they felt sad and to indicate, on a 5-point Likert scale ranging from 0 (i.e. 'not at all') to 4 ('very strongly'), the degree to which each statement of the 34 items applied to them. The LEIDS-r total score was then calculated by adding the scores of its six sub-scales: aggression, control, hopelessness, risk aversion, rumination and acceptance.

BIS-BAS: We assessed the affective response (sensitivity) to reward and punishment using the Behavioral Inhibition Scale (BIS) / Behavioral Approach Scale (BAS) (Carver and White 1994).

Table S1 shows the scores on the questionnaires before and after the intervention for the placebo and probiotics group

Table S1. *Raw scores on the questionnaires expressed as mean values (SEM).*

	PLACEBO pre	**PLACEBO post**	**PROBIOTICS pre**	**PROBIOTICS post**	Group*Session p-values
BDI	2.62 (0.54)	2.24 (0.58)	2.10 (0.37)	1.78 (0.39)	.95
LEIDS-r	28.59 (2.58)	21.66 (2.08)	29.97 (2.35)	27.24 (2.55)	.09
BIS	21.41 (0.68)	21.38 (0.60)	20.45 (0.78)	19.83 (0.69)	.32
BAS-Reward	16.72 (0.32)	16.79 (0.28)	16.38 (0.34)	16.31 (0.36)	.76
BAS-Fun Seeking	11.55 (0.30)	11.76 (0.30)	12.03 (0.33)	11.79 (0.43)	.34
BAS-Drive	13.21 (0.34)	13.21 (0.35)	12.93 (0.28)	12.97 (0.35)	.93

DIET: furthermore, at both sessions, participants were asked to indicate any change in diet type (options: restrictive, low-sodium, high-fiber, high-cholesterol, anti-diabetes diet) and diet style (options: vegetarian, vegan, macrobiotic, anthroposophic diet, no meat, non-specified). Participants with very different or specific diets were excluded *a priori* during the first phone screening.

Moreover, participants completed a short Food Frequency Questionnaire, called the Dutch Healthy Diet index (FFQ-DHD). The FFQ-DHD assesses the degree to which participants eat according to the national guidelines for a Dutch healthy diet (max score is 80; 10 per component) based on the following components: vegetables, fruit, fibre, fish, saturated fat, trans fat, salt and alcohol (van Lee, Feskens et al. 2013).

Table S2 shows the diet type, diet style, and FFQ-DHD scores within the last month before the start of the supplementation with probiotics or placebo and during the month of intervention.

Table S2. *Diet-related information pre- and post-intervention. For diet type and diet style, only a change relative to pre-intervention is indicated at post-intervention. For the FFQ-DHD, the average score (SEM) is given for n=54 pre-intervention and n=52 post-intervention.*

	PLACEBO pre	**PLACEBO post**	**PROBIOTICS pre**	**PROBIOTICS post**
Diet type	*Regular: 29*		*Regular: 29*	
(N particip.)	*Restrictive: 0*		*Restrictive: 0*	
	Low-sodium: 0		*Low-sodium: 0*	
	High fiber: 0		*High fiber: 0*	
	High cholesterol: 0		*High cholesterol: 0*	
	Anti-diabetes: 0		*Anti-diabetes: 0*	
	No-specified: 0	*No-specified: 1*	*No-specified: 0*	*No-specified: 1*
Diet style	*Regular: 28*		*Regular: 24*	*Regular:1*
(N particip.)	*Vegetarian: 1*		*Vegetarian: 2*	
	Vegan: 0		*Vegan: 0*	
	Macrobiotic: 0		*Macrobiotic: 0*	
	Anthroposophic: 0		*Anthroposophic: 0*	
	No-specified: 0	*No-specified: 1*	*No-specified: 3*	
FFQ-DHD - total	*52.4 (1.8)*	*54.5 (2.2)*	*48.4 (2.2)*	*51.5 (2.1)*
vegetables	*6.2 (0.5)*	*6.1 (0.5)*	*5.8 (0.5)*	*5.8 (0.5)*
fruit	*7.8 (0.5)*	*8.1 (0.5)*	*7.8 (0.4)*	*7.0 (0.7)*
fibre	*7.5 (0.4)*	*7.7 (0.3)*	*7.8 (0.4)*	*7.8 (0.4)*
fish	*4.4 (0.5)*	*4.8 (0.6)*	*4.1 (0.6)*	*4.5 (0.6)*
saturated fat	*4.1 (0.7)*	*5.3 (0.8)*	*4.0 (0.8)*	*6.3 (0.8)*
trans fat	*8.2 (0.8)*	*7.8 (0.8)*	*6.3 (0.9)*	*7.2 (0.9)*
salt	*7.3 (0.4)*	*6.8 (0.5)*	*6.2 (0.6)*	*6.6 (0.5)*
alcohol	*6.9 (0.8)*	*8.0 (0.7)*	*6.4 (0.9)*	*6.3 (0.9)*

Note. FFQ-DHD = Food Frequency Questionnaire – Dutch Healthy Diet index

Analysis using the Exact F of Fisher test showed no differences in specific diet type (F Exact p = 1.000), or in diet-related styles between the two groups at baseline (F Exact p = 1.000). No differences were found comparing the two groups pre- versus post-treatment in specific diet type (F Exact p = 0.670), or diet style (F Exact p = 0.194).

For the FFQ-DHD analysis, we had missing information for n=2 in the placebo group and n=2 in the probiotics group at both the pre- and post-intervention session. An additionally n=2 datasets were missing in the probiotics group post-intervention. At baseline (placebo: n=27; probiotics: n=27), we did not find any group differences in the FFQ-DHD total score or its sub-components (all: $-1.5 < t(52) < 1.7$, $p>.1$). We also did not find any differences when comparing the two groups pre- versus post-treatment (placebo: n=27; probiotics: n=25; Group * Session, all $F(1,50)<2.5$, $p>.1$).

Preprocessing fMRI data and first level analyses

Volumes for each echo-time were realigned using six rigid body spatial transformations (translations and rotations: x, y, z, pitch, roll, jaw). Thirty volumes acquired before the tasks were used to combine the four echo images into a single volume using an echo weighting method known as PAID-weighting (Poser, Versluis et al. 2006). Resulting combined functional (EPI) images were slice-time corrected by realigning the time series for each voxel to the time of acquisition of the reference slice. Subject-specific structural and functional data were subsequently co-registered to a standard structural or functional stereotactic space respectively, using Montreal Neurological Institute (MNI) templates. A unified segmentation approach was then used to segment the structural images. Segmented images were subsequently spatially co-registered to the mean of the functional images. The transformation matrix resulting from the segmentation step was used to normalize the structural and functional images to MNI space,

resampled at a voxel size of 2 x 2 x 2 mm. In a final step, normalized functional images were spatially smoothed using an 8 mm full-width at half maximum (FWHM) Gaussian kernel.

Fixed effects analyses of the Emotional face-matching paradigm were carried out at the first level using a block-design fMRI approach (i.e. 12 'emotion' blocks and 6 'shape' blocks of each 17 seconds). Two regressors of interest were compared: 'emotion' minus 'shape' condition. Similar first level analyses were performed using an event-related approach for the Stroop paradigms. For both the emotional face-word Stroop paradigm and the color-word Stroop paradigm the statistical model contained two regressors of interest, which we subtracted for our contrast of interest: 'incongruent' minus 'congruent' condition. Missed and incorrect trials were taken into account in a regressor of non-interest for both paradigms. Additionally, thirteen regressors of non-interest were added to the designs of all three tasks, including twelve rigid-body transformation parameters (i.e. movement regressors consisting of three translations, three rotations and their linear derivatives) obtained during realignment, as well as one constant term. A high-pass filter with a cut-off of 128 seconds was applied to the time-series of the functional images to remove low-frequency drifts. By applying an autoregressive AR (1) model, correction for serial correlations was carried out. Both sessions of each subject were included in one first level model.

Reaction Times on fMRI tasks

The analyses were done on log-transformed data to reduce the skewness of the distributions, which were better normalized after this correction. In the emotional face-matching task, participants were significantly slower in the 'emotion' than in the 'shape' condition (main Condition: $F(1,56) = 654.65$, $p < .001$, $\eta^{p2}=.921$). Similarly, during the two Stroop tasks, participants were significantly slower in the incongruent than in the congruent conditions (emotional face-word Stroop, main Condition: $F(1,56) =$

276.60, p<.001 , η^{p2}=.921; color-word Stroop, main Condition: F(1,56) = 231.05, p<.001 , η^{p2}=.805). See **Table S3** for the raw response times per task.

Table S3. *Mean (SEM) RTs for each fMRI paradigm, before and after the intervention with probiotics and placebo.*

Paradigm	**Condition**	PLACEBO pre	PLACEBO post	PROBIOTICS pre	PROBIOTICS post
Emotional face-matching	Shape	867.85 (36.9)	775.18 (27.29)	821.46 (22.50)	779.03 (25.89)
	Emotion	1352.73 (57.67)	1282.57 (54.56)	1364.69 (68.81)	1253.71 (59.53)
Emotional face-word Stroop	Congruent	778.44 (29.85)	795.55 (37.72)	761.60 (41.50)	762.11 (33.39)
	Incongruent	829.63 (30.63)	847.29 (39.55)	825.89 (44.18)	812.06 (33.01)
Color-word Stroop	Congruent	856.71 (42.21)	817.61 (35.09)	801.45 (52.49)	784.00 (38.22)
	Incongruent	974.80 (44.67)	955.49 (43.02)	917.32 (52.07)	893.74 (51.26)

No significant Group x Session x Condition effects

fMRI results

Emotion reactivity was seen for instance in the amygdala and in terms of RTs during emotional face matching (Hariri, Bookheimer et al. 2000, Haxby, Hoffman et al. 2000), while emotion regulation and

general cognitive control was observed in frontal regions and in longer RTs during the two Stroop tasks (Etkin, Egner et al. 2006, Aarts, Roelofs et al. 2008, Roberts and Hall 2008, Aarts, Roelofs et al. 2009, Cieslik, Mueller et al. 2015). See **Table S4** for the activated clusters at $p_{FWE}<.05$.

Table S4. *Main task activations at $P_{FWE}<.05$ (cluster level) across sessions and groups.*

Emotional face-matching	*Size (# voxels)*	*MNI coordinates*			***Emotional face-word Stroop***	*Size (# voxels)*	*MNI coordinates*			***Color-word Stroop***	*Size (# voxels)*	*MNI coordinates*		
Emotion > shape		*x*	*y*	*z*	***Incongruent > congruent***		*x*	*y*	*z*	***Incongruent > congruent***		*x*	*y*	*z*
L Inf Occ cortex	*50411*	*-24*	*-94*	*-8*	L Intrapar s	*12607*	*-30*	*-52*	*48*	L lat PFC	*24074*	*-42*	*20*	*24*
L Thal / Amygdala		*-22*	*-28*	*-4*	R Fusiform g	*3174*	*40*	*-48*	*-18*	L pre-SMA		*-6*	*2*	*60*
R Inf Front g		*46*	*22*	*24*	R Intrapar s	*3380*	*28*	*-58*	*46*	L Occ Temp cortex	*1633*	*-48*	*-56*	*-8*
R Thal / Amygdala		*22*	*-28*	*-2*	L pre-SMA	*4635*	*-6*	*8*	*54*	R Cerebellum	*1657*	*12*	*-74*	*-24*
L Intrapar s	*916*	*-28*	*-58*	*46*	*Congruent > incongruent*		*x*	*y*	*x*	R lat PFC	*698*	*48*	*6*	*30*
R Intrapar s	*595*	*30*	*-54*	*46*	L Angular g	*727*	*-42*	*-76*	*36*	L Sup Temp g	*382*	*-50*	*-48*	*12*
L Sup Temp g	*571*	*-46*	*-46*	*12*	L vmPFC	*4448*	*-6*	*52*	*-6*					
Shape > emotion		*x*	*y*	*z*	L Sup Front g	*988*	*-22*	*24*	*46*					
L rostr ACC	*4783*	*-8*	*36*	*2*	L Mid Temp g	*756*	*-60*	*-4*	*-16*					
L Sup Temp s	*297*	*-58*	*-64*	*6*	L Post Cing cortex	*1093*	*-10*	*-44*	*36*					
L Parahip cortex	*491*	*-38*	*-80*	*38*	R Mid Temp g	*160*	*60*	*-2*	*-14*					
R Angular g	*962*	*50*	*-64*	*32*										
L Hippocamp	*213*	*-20*	*-42*	*10*										
R Sup Temp g	*1131*	*62*	*-24*	*-10*										
R Post Cing cortex	*356*	*2*	*-26*	*40*										
L Mid Temp g	*304*	*-64*	*-26*	*-18*										
L Sup Front g	*448*	*-20*	*30*	*38*										
L Sup Temp g	*303*	*-46*	*-38*	*20*										

Note. L = left; Inf = Inferior; Occ = occipital; Thal = thalamus; R = right; Front = frontal; g = gyrus; Intrapar = intraparietal; s = sulcus; Sup = superior; Temp = temporal; rostr = rostral; ACC = anterior cingulate cortex; Parahip = parahippocampal; Hippocamp = hippocampus; Cing = cingulate; Mid = middle; pre-SMA = pre-supplementary motor area; vmPFC = ventromedial prefrontal cortex; lat PFC = lateral PFC

Saliva and cardiovascular data

Participants were instructed not to consume food and beverages other than water or to exercise within the two hours preceding the start of the saliva collection. The saliva samples were collected via absorbent devices (salivettes -Sarstedt, Nümbrecht, Germany) and immediately frozen (at the temperature of -24°). Cortisol and alpha-amylase parameters were analyzed at the end of the experiment by an independent and specialized lab. Both sessions were conducted in the afternoon (after 13:00 PM).

Cardiovascular monitoring, i.e. heart rate and blood pressure (systolic -BPsys- and diastolic - BPdia -), were assessed using a standard upper arm blood pressure monitor medical device.

Stress-related results

For the cardiovascular parameters BPsys and BPdia, we found significant effects of the stressor (main Time (7), BPsys: $F(6,51) = 20.7$, $p < .001$, $\eta^{p2} = .708$; BPdia: $F(6,51) = 8.3$, $p < .001$, $\eta^{p2} = .493$) and of repeating the stressor (Session(2), BPsys: $F(1,56) = 6.01$, $p = .01$, $\eta^{p2} = .097$; BPdia: $F(1,56) = 5.7$, $p = .02$, $\eta^{p2} = .093$). For both blood pressures we did not find a significant Time(7) x Group(2) x Session(2) interaction (BPsys: $F(6,51)<1$, $\eta^{p2} = .086$ and BPdia: $F(6,51)<1$, $\eta^{p2} = .091$), or any other significant interaction (all $p>.05$). For HR, cortisol, and alpha-amylase levels, we similarly observed an effect of the stressor (main Time (7) HR: $F(6,51) = 21.9$, $p < .001$, $\eta^{p2} = .721$; main Time (5) cortisol: $F(4,48) = 16.8$, p

<.001 , η^{p2}=.583; main Time (5) alpha-amylase: F(4,36) = 3.9 , p =.01, η^{p2}=.304), but no other main or interaction effects (all p>.05).

We calculated a total VAS score by adding up the six subscales (irritation, tension, happiness (reverse scoring), pain, fear, and stress levels) to obtain one measure of subjective feeling of stress. We found a significant effect of the stressor (Time(5): F(4,52) = 36.8, p <.001 η^{p2}=.739) and of repeating it (Session(2): F(1,55) = 4.7, p = .035, η^{p2}=.078), and a significant interaction between Time and Session (Time*Session: F(4,52) = 2.7, p = .04, η^{p2}=.171); however, these effects did not differ across groups (all interactions with Group, p>.05). We also did not find differences between the probiotics and placebo group at baseline, i.e. pre-intervention, for any of the physiological or subjective stress variables (all p>.05) except for the sub-scale VAS 'tension' (p =.02) where the probiotics group indicated to feel more tense (SD): mean 2.79 (1.8), in comparison with the placebo group (SD): mean 1.93 (.99). No group differences at baseline were observed for the total VAS score. To sum up, although the stressor worked, the absence of interaction with Session and Group demonstrated that physiological and subjective stress measures were not affected by the probiotics.

Stress-induced working memory performance

As expected, the effects were specific to the cognitively more demanding digit span backward test (see main text), as we did not find a significant Time(2) x Group(2) x Session(2) interaction for digit span forward performance (F(1,56)<1). For raw digit span scores, see **Table S5** below.

Table S5. *Mean (SEM) Digit Span scores (forward and backward) before and after stress induction, and before and after the supplementation period with placebo or probiotics.*

	PLACEBO pre	PLACEBO post	PROBIOTICS pre	PROBIOTICS post
DS forward				
before SECPT	8.00 (0.36)	8.10 (0.34)	8.55 (0.43)	8.93 (0.40)
DS forward				
after SECPT	7.62 (0.35)	8.72(0.31	8.89 (0.44)	9.82 (0.48)
DS backward				
before SECPT	6.83 (0.32)	7.90 (0.32)	8.17 (0.40)	8.48 (0.40)
DS backward				
after SECPT	7.14 (0.34)	7.86 (0.34)	7.76 (0.41)	8.90 (0.48)

Correlations between neural Stroop responses and stress-induced working memory effects: whole-brain analysis

We used the post- minus pre-intervention stress-related working memory scores as a regressor of interest in the original t-test model of the fMRI data of the color-word Stroop task (i.e. incongruent > congruent, post- > pre-intervention), separately for each group. Only within the probiotics group, we found whole-brain corrected significant associations between probiotic-induced (post-pre intervention) increases in stress-related DS backwards and probiotic-induced decreases in brain responses (incongruent-congruent) during the color-word Stroop task in striatum, bilateral PFC, and medial frontal cortex (p_{FWE}<.05 at cluster level, **Table S6** + **Figure S1)**.

Table S6. *Whole-brain corrected significant clusters (p_{FWE}<.05, cluster defining threshold: p<.001) activated during the color-word Stroop task (post-pre intervention) that were correlated with stress-related changes in working memory in the probiotics group (post-pre intervention).*

	Size (# voxels)	**MNI (x, y, z)**	**$p_{FWE\text{-}cor}$ at cluster level**
R lat PFC	967	40, 12, 20	< 0.001
R Caudate nucleus	1275	6, 4, 8	< 0.001
R frontopolar cortex	232	22, 64, 10	0.001
pre-SMA	558	4, 18, 54	< 0.001
L lat PFC	468	-56, 16, 22	< 0.001

Note. lat PFC = lateral prefrontal cortex; pre-SMA = pre-supplementary motor area

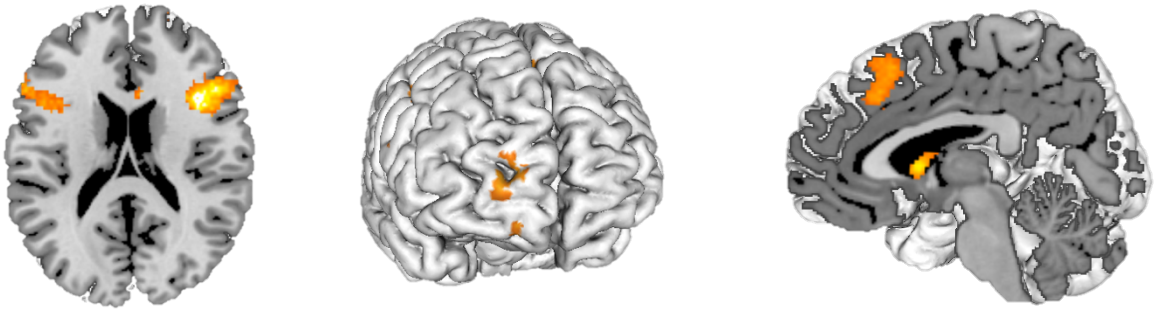


Figure S1. Negative correlation between stress-related DS backwards performance (post- minus pre-intervention) and incongruent versus congruent responses during the color-word Stroop task (post- minus pre-intervention) in the probiotics group. Only significant clusters are shown (whole-brain p_{FWE}<.05, cluster defining threshold: p<.001).

Supplemental Bibliography

Aarts, E., A. Roelofs and M. van Turennout (2008). "Anticipatory activity in anterior cingulate cortex can be independent of conflict and error likelihood." J Neurosci **28**(18): 4671-4678.

Aarts, E., A. Roelofs and M. van Turennout (2009). "Attentional control of task and response in lateral and medial frontal cortex: brain activity and reaction time distributions." Neuropsychologia **47**(10): 2089-2099.

Carver, C. S. and T. L. White (1994). "Behavioral-Inhibition, Behavioral Activation, and Affective Responses to Impending Reward and Punishment - the Bis Bas Scales." Journal of Personality and Social Psychology **67**(2): 319-333.

Cieslik, E. C., V. I. Mueller, C. R. Eickhoff, R. Langner and S. B. Eickhoff (2015). "Three key regions for supervisory attentional control: evidence from neuroimaging meta-analyses." Neurosci Biobehav Rev **48**: 22-34.

Etkin, A., T. Egner, D. M. Peraza, E. R. Kandel and J. Hirsch (2006). "Resolving emotional conflict: a role for the rostral anterior cingulate cortex in modulating activity in the amygdala." Neuron **51**(6): 871-882.

Hariri, A. R., S. Y. Bookheimer and J. C. Mazziotta (2000). "Modulating emotional responses: effects of a neocortical network on the limbic system." Neuroreport **11**(1): 43-48.

Haxby, J. V., E. A. Hoffman and M. I. Gobbini (2000). "The distributed human neural system for face perception." Trends Cogn Sci **4**(6): 223-233.

Poser, B. A., M. J. Versluis, J. M. Hoogduin and D. G. Norris (2006). "BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: parallel-acquired inhomogeneity-desensitized fMRI." Magn Reson Med **55**(6): 1227-1235.

Roberts, K. L. and D. A. Hall (2008). "Examining a supramodal network for conflict processing: a systematic review and novel functional magnetic resonance imaging data for related visual and auditory stroop tasks." J Cogn Neurosci **20**(6): 1063-1078.

Steenbergen, L., R. Sellaro, S. van Hemert, J. A. Bosch and L. S. Colzato (2015). "A randomized controlled trial to test the effect of multispecies probiotics on cognitive reactivity to sad mood." Brain Behav Immun **48**: 258-264.

van Lee, L., E. J. Feskens, E. J. Hooft van Huysduynen, J. H. de Vries, P. van 't Veer and A. Geelen (2013). "The Dutch Healthy Diet index as assessed by 24 h recalls and FFQ: associations with biomarkers from a cross-sectional study." J Nutr Sci **2**: e40.