PNAS

Supplementary Information for
Expectation Effects in Working Memory Training

Jocelyn Parong, Aaron R. Seitz, Susanne M. Jaeggi, C. Shawn Green

Corresponding authors. Jocelyn Parong; C. Shawn Green
Email: jparong@wisc.edu; cshawn.green@wisc.edu

This PDF file includes:

Supplementary text
Figures S1 to S4 (not allowed for Brief Reports)
Tables S1 to S3 (not allowed for Brief Reports)
SI References

Supplementary Information Text

Participant Exclusions

Figure S1 shows the number of participants who were excluded at each step of the study. It should be noted that those who did not comply with training directions were still able to participate in the post-test and delayed-test, but were not included in the analyses.

Expectation Induction

Participants were presented with a short explanation of why the cognitive training is expected to improve or decrease their cognitive performance, which included exaggerated or false scientific studies to support these expectations. We note that the only completely false study is the "Switzerland study" in the negative expectation condition, while other studies were exaggerated descriptions of real results. Participants were made aware of this during debriefing. Participants were provided with the following text according to their assigned expectation condition:

- Positive Expectation: You will be playing a game that targets specific components of cognition. You will repeatedly practice using these cognitive skills, so you can expect to improve in the game. Additionally, previous research has shown that your performance on cognitive tasks outside of the game, including on the same lab tasks that you've just completed, will also improve. This effect is similar to when you practice playing one song on the piano and you get better at playing a new song on the piano. Or, if you practice memorizing letters, you also get better at memorizing numbers.

  Evidence for these benefits can also be seen in the brain. The brain is malleable, meaning it can change and form new neural connections, making certain areas of the brain stronger. This is important because it seems that the cognitive functions associated with specific brain regions can also be improved. Cognitive training programs cause an increase in neural connections to the parts of the brain that are important in cognition. The parts of the brain that are involved in cognition are very connected, so an increase of neural connections in one area also increases neural connections in another area.

  Previous research has shown that training programs like the one you will play have improved people's cognitive abilities. For example, in one study, participants played a video game called Neuroracer. They improved their cognitive control abilities, including multitasking, attention, and working memory. Additionally, EEG showed increased activity in brain areas responsible for those parts of cognition, which are also involved in other parts of cognition. In another study, researchers found that cognitive training improved performance on working memory tasks, and fMRI scans showed increased brain connectivity in the fronto-parietal network, which plays a crucial role in cognitive function.
- Negative Expectation: You will be playing a game that targets specific components of cognition. You will repeatedly practice using these cognitive skills, so you can expect to improve in the game. However, there is no evidence that cognitive training improves performance on cognitive tasks outside of the training environment, like the tasks you just completed in the lab. In fact, because you will practice only the skills required in the game, you may actually see a decrease in your performance on the lab tasks. Critically, this does NOT mean that the game decreases your general cognitive abilities. This effect is similar to getting very used to driving one particular car, and then switching to a new car and feeling like it is difficult. Or, if you've only typed on your own laptop keyboard, it can be hard to switch to a new keyboard.

  Evidence for this can also be seen in the brain. The brain is malleable, meaning it can change and form new neural connections, making certain areas of the brain stronger. However, brain areas involved in cognition are quite distinct from one another, so increasing neural connections in one area does not increase neural connections in another. Cognitive training programs cause an increase in neural connections only in the area that they target, but not other areas.

  Previous research has shown that training programs like the one you will play do not improve people's cognitive abilities. For example, in one study called "Cognitive training does not enhance general cognition," the researchers reviewed many other experiments on cognitive training programs, including working memory training and video game training, and found that there is no evidence that they were effective. In another review, Simons et al. (2016) concluded that cognitive training programs used in clinical trials and commercial brain-training programs, including the brain-training game Lumosity, were not effective at improving everyday cognitive performance. There were also no significant changes in brain areas associated with cognitive performance.

A recent study in Switzerland has shown evidence that cognitive training actually decreases performance in other tasks that were not related to the training game. Participants who played a working memory game performed worse on non-working memory tasks. Additionally, the researchers found less activity in some regions that were not associated with working memory, including the posterior temporal gyrus and dorso-lateral posterior frontal gyrus, which are important in other cognitive abilities.

Mid-Test Difficulty Detection

Previous pilot work suggested that cognitive task difficulty can be manipulated without detection. Participants were given 3 blocks of the fluid intelligence, mental rotation, tasking-switching, ANT, and UFOV tasks. The first block was always a standard version (i.e., versions that will be utilized as pre-test or post-test assessments in the proposed studies), as was one of the following two blocks. The remaining block of the three was the easier version, like the tasks used in the placebo mid-test here. Several outcomes were noted. First, and foremost, was whether the "easier version" was in fact easier (i.e., whether performance was significantly greater in the mid-test version). Second, was whether participants could identify in which block they performed best (they were asked to rank their performance in the three blocks). Third was whether participants could indicate, for the block that they selected as the easiest, why they performed best in that block. The results of this preliminary data are summarized in Table S1.

Mid-test Manipulations

The cognitive battery was manipulated for the mid-test, such that those who received the placebo completed a mid-test that was easier than the pre-test and those who received the nocebo completed a mid-test that was more difficult than the pre-test. Critically, the tests were manipulated in a way that was intended to not be obvious that they were manipulated. Manipulations for the specific tasks are outlined below. Additionally, comparisons of performance between the two tests are shown in Table S2.

O-Span*.* Task-difficulty was manipulated in two ways. First, in the placebo test, the interleaved math problems were made easier by using only numbers less than 5 (e.g., [(2*3) -1 =?]), while in the nocebo test, these were made more difficult by using only numbers greater than 5 (e.g., [(7*8)-9]). Second, the placebo test contained a higher proportion of shorter letter sets (i.e., 3 and 4 letter sets), while the nocebo test contained a higher proportion of longer letter sets (i.e., 6 and 7 letter sets).

Task-switching*.* In the placebo test, there was a higher proportion of neutral trials (i.e., letter and number trials paired with a symbol rather than a number or letter). In the nocebo test, there was a higher proportion of incongruent trials (i.e., letter-number pairs in which the response keys for the letter and number response were not the same).

Countermanding. In the placebo test, there was a lower proportion of incongruent trials (pressing the key on the opposite side of the stimulus) and a lower proportion of switch trials compared to the pre-test, while in the nocebo test, there were higher proportions of both incongruent and switch trials compared to the pre-test.

UFOV. In the placebo test, the presentation duration of trials was increased, while in the nocebo test, the presentation duration of trials was decreased.

ANT. The placebo test was made easier by increasing the proportion of neutral trials (i.e., trials with no distracting flanker stimuli) and decreasing the number of incompatible trials (i.e., trials in which the flanker stimuli were the opposite of the target stimulus), while conversely the nocebo test was made more difficult by decreasing the proportion of neutral trials and increasing incompatible trials.

Mental Rotation. In the placebo test, relative to the pre-test, there was a higher proportion of trials with small angles of rotation (e.g., 10-90 degrees), which decreased difficulty, while in the nocebo test, relative to the pre-test, there was a higher proportion of trials with large angles of rotation (e.g., 110-190 degrees), which increased difficulty.

Paper Folding. The placebo test had a higher proportion of more difficult trials, which contained more or more complicated (diagonal/partial) folds, while the nocebo test had a higher proportion of easier trials, relative to the proportion of easy/hard trials on the pre-test.

RAPM and UCMRT. In both of these tasks, the matrix problems had known levels of difficulty. Thus, the placebo test contained a higher proportion of easier items, while the nocebo test contained a higher proportion of more difficult items, relative to the proportion of easy/hard items in the pre-test.

Mill-Hill and Shipley vocabulary. In both of these tasks, items had known levels of difficulty. Thus, the placebo test contained a higher proportion of easier items, while the nocebo test contained a higher proportion of more difficult items, relative to the proportion of easy/hard items in the pre-test.

Researcher (Un)awareness

In an attempt to minimize experimenter effects from influencing participants' performance on the cognitive battery (e.g., unintentional encouragement after knowing a participant is in the placebo condition), researchers (R1) assigned to all pre-test and post-test sessions were intentionally not informed of the expectation manipulations in the study. However, to properly instruct participants about certain aspects of the study, some researchers could not remain unaware to their conditions; researchers (R2) were aware of the full study design and assigned to administer the expectation induction, instructions about the training game, mid-test, debriefing, and delayed tests (as participants were aware of their expectations at this point). After each testing session in which researchers interacted with participants, researchers were asked to guess A) which training condition the participant was in among 3 choices (experimental, control, or other) and B) how the participant expected their cognitive performance to change among 3 choices (improve, get worse, or stay the same). Thus, it was predicted that R1s should not correctly guess participant training or expectation conditions above chance levels (.33). To test this, chi-squares were run on perceptions of training condition and cognitive improvement across after the pre-test and post-test. There were 15 missing data points. As shown in Table S3, R1s were not able to correctly guess participants' expectations above chance levels, X^2 (1, N = 235) = 0.42, p = .518. R1s also guessed participants' training condition significantly *below* chance levels, X^2 (1, N = 235) = 7.16, p = .007.

Control analyses

Two 2 (positive vs. nocebo) x 2 (true cognitive training vs. control training) MANCOVAs were run on the post-test and delayed-test task performance on the vocabulary measures (Mill Hill, Shipley), with respective pre-test scores as covariates. At post-test, the expectation effect, F(2,118) = 0.43, p = .649, ηp2 = 0.01, training effect, F(2,118) = 0.93, p = .398, ηp2 = 0.02, and interaction between the two, F(2,118) = 0.00, p = .996, ηp2 = 0.00, on overall performance were not significant. At delayed-test, the effects remained the same; the expectation effect, F(2,118) = 0.65, p = .523, ηp2 = 0.01, training effect, F(2,118) = 0.20, p = .823, ηp2 = 0.00, and interaction between the two, F(2,118) = 1.16, p = .316, ηp2 = 0.02, on overall performance were not significant.

Additional moderator analyses

Additional exploratory moderator analyses examined interactions between the expectation effect and a number of variables on specific post-test task performance, with respective pre-test scores as covariates. The moderators examined included gender, age, subscale scores of the Big Five Personality Inventory (openness, conscientiousness, extraversion, agreeableness, and neuroticism), motivation subscale scores from the Behavioral Inhibition System and Behavioral Activation System Scales (BIS/BAS; drive, fun-seeking, reward responsiveness, and BIS total), grist scale score, metacognitive scale score, fixed/growth mindset scale score, subscale scores of the Schutte Self-Report Emotional Intelligence Test (emotion perception, utilizing emotions, managing self-relevant emotions, and managing others emotions), subscale scores of the Work and Family Orientation scale (hard work, mastery, and competitiveness). Simple moderations were conducted using the PROCESS macro for SPSS, using 5000 bootstrap samples for bias correction. Table S4 and Figure S2 show the interaction effect between each moderator and the expectation manipulation on each task. A few significant patterns emerged. Mindset and utility of emotions (SSEIT subscale) significantly moderated the effect of expectations on the O-span. Metacognition, perception of emotions (SSEIT subscale), and work (WOFO subscale) significantly moderated the effect of expectations on the mental rotation task. The BIS significantly moderated the effect of expectations on ANT performance.

Far Transfer Mediation

Our main results replicated well-established near transfer effects of n-back working memory training to an untrained n-back, and a small (but non-significant) effect on far transfer training effects to fluid intelligence measures. To further test a mediation model that suggests far transfer occurs through near transfer gains (Pahor et al., 2022), performance on the untrained n-back task in the post-test was tested as a mediator. As a control analysis, performance on the countermanding task, as used in Pahor et al., was entered as a parallel mediator, with pre-test n-back, countermanding, and matrix scores as covariates. A composite fluid intelligence metric was calculated from the average z-scored gain on the UCMRT and RAPM. In the mediation analysis (Figure S3), post-test n-back and countermanding scores were added as parallel mediators. Critically, the indirect effect of training group on post-test matrix reasoning performance through n-back performance was significant, $b = -0.18$, $SE = .07$, 95% CI [-.34, -.05], while the indirect effect through countermanding performance was not significant, $b = 0.02$, $SE = .02$, 95% CI [-.03, .06]. Further, these two indirect effects significantly differed from each other, $b = -0.20$, $SE = .07$, 95% CI [-.36, -.05]. Thus, these results replicate previous findings.

With respect to the expectation manipulations used in this study, a moderated mediation model was tested to determine whether these indirect effects differed by expectation condition. Expectation was entered as a moderator variable along all the paths from the training condition to the mediators and outcome variable. Expectation did not significantly moderate the effect of training group on the n-back, $F(1,121) = 0.22$, $p = .634$, countermanding, $F(1,124) = 0.16$, $p = .690$, or matrix test performance, $F(1,119) = 0.14$, $p = .706$. However, in examining the conditional indirect effects, as shown in Figure S4, the training effect on post-test matrix reasoning performance through n-back performance seemed to be stronger for the positive expectation group, -0.21, $SE = .09$, 95% CI [-.41, -.04], compared to the negative expectation group, $b = -0.15$, $SE = .10$, 95% CI [-.37, .03]. However, these were not significantly different, $b = 0.06$, $SE = .13$, 95% CI [-.20, .32]. Both the training effect on post-test matrix reasoning performance through countermanding performance were not significant for the placebo group, $b = 0.03$, $SE = .03$, 95% CI [-.03, .10], and nocebo group, $b = 0.02$, $SE = .03$, 95% CI [-.04, .09], and these effects were not significantly different from each other, $b = -.01$, $SE = .04$, 95% CI [-.10, .06].

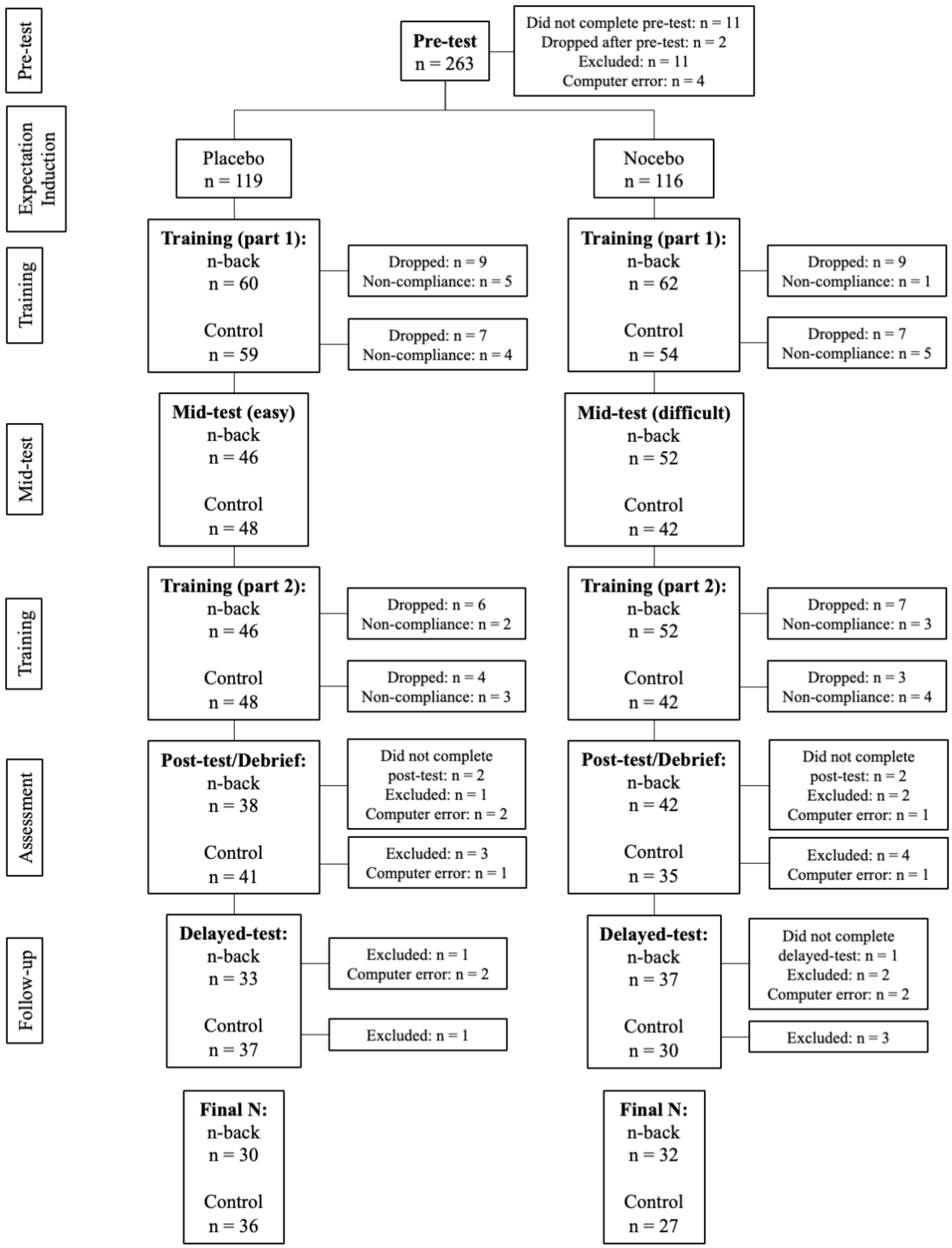

Figure S1. Participant attrition during the study.

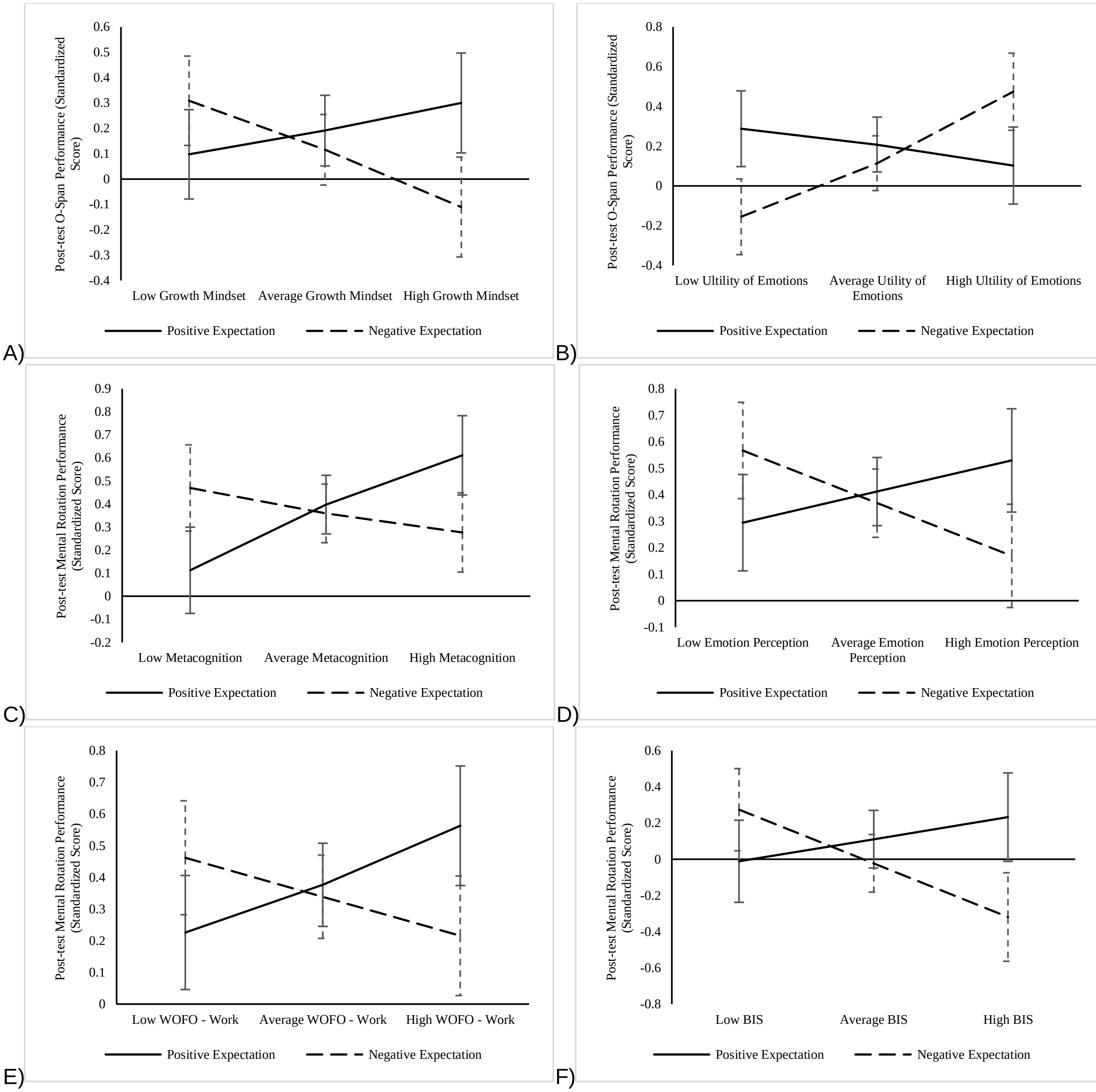

Figure S2.

Significant moderator effects for A) mindset and the expectation effect on O-span performance, B) SSEIT - utility subscale and the expectation effect on O-span performance, C) metacognition and the expectation effect on mental rotation performance, D) SSEIT - perception subscale and the expectation effect on mental rotation performance, E) WOFO - work subscale and the expectation effect on mental rotation performance, and F) BIS and the expectation effect on ANT performance. All moderation analyses controlled for respective pre-test task performance.

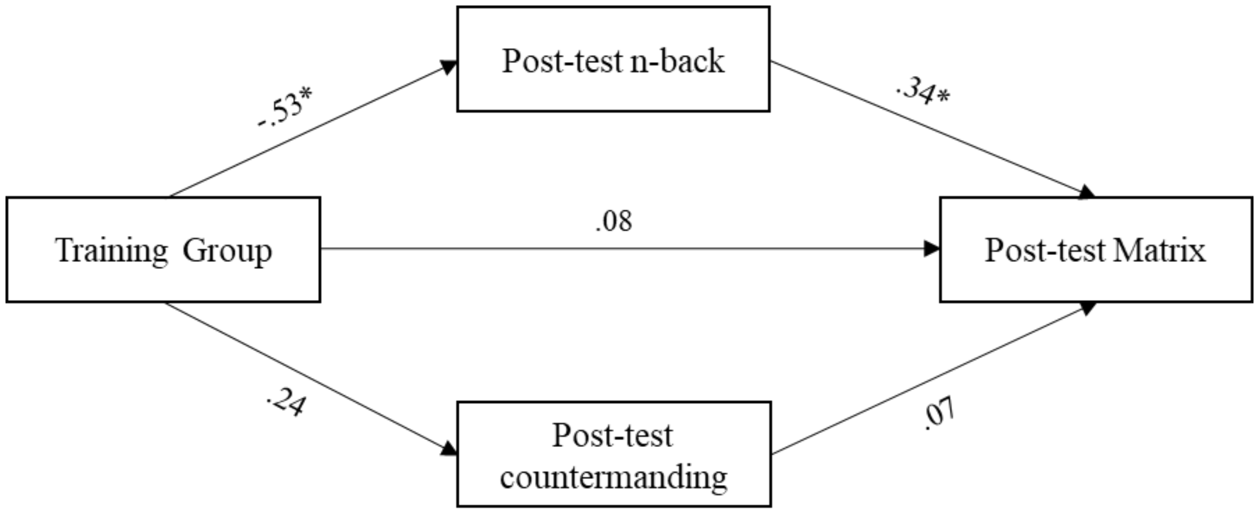

Figure S3.
Unstandardized regression coefficients for the relationship between training group (categorical; 1 = n-back, 2 = trivia) and performance on post-test matrix reasoning tasks (UCMRT and RAPM) as mediated by post-test n-back and post-test countermanding performance. A negative training effect on n-back performance indicates better performance for the n-back training group.The model includes pre-test scores (n-back, countermanding, matrix) as covariates.

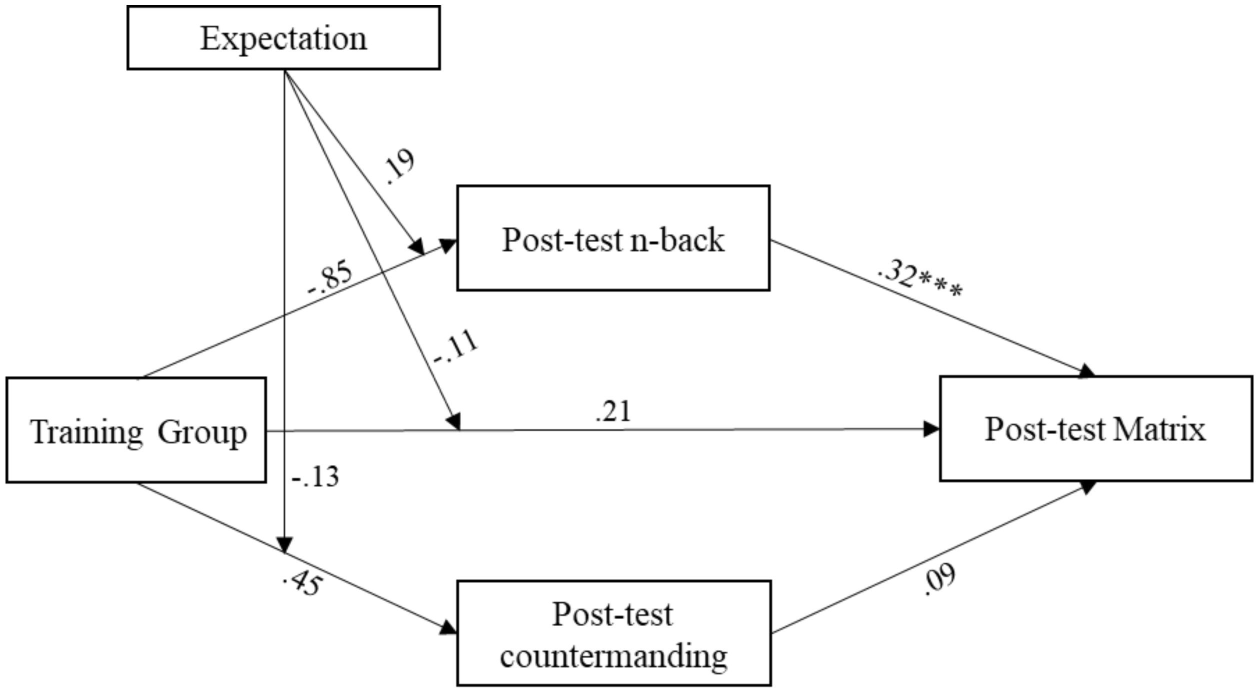

Figure S4.
Unstandardized regression coefficients for the relationship between training group (categorical; 1 = n-back, 2 = trivia) and performance on post-test matrix reasoning tasks (UCMRT and RAPM) as mediated by post-test n-back and post-test countermanding performance and moderated by expectation condition (categorical; 1 = placebo, 2 = nocebo). A negative training effect on n-back performance indicates better performance for the n-back training group. The model includes pre-test scores (n-back, countermanding, matrix) as covariates.

Table S1
Pilot results of easy mid-test task difficulty manipulation detection.

	N	# that performed best in mid-test version	# who could identify the block they performed best in	# who indicated explicit (correct) awareness of why the mid-test version was easiest
Fluid Intelligence	15	15	15	1
Mental Rotation	13	13	13	0
Task-Switching	12	11	9	0
UFOV	14	14	13	2
ANT	14	12	10	0

Table S2
Performance on the mid-test by expectation condition.

Task	placebo mid-test (easy) N = 66	nocebo mid-test (difficult) N = 59	t-test (one-tailed)
O-span (total score)	M = 27.72, SD = 9.94	M = 15.76, SD = 12.16	t(122) = 6.02, p < .001
Task-switching (IES)	M = 855.93, SD = 302.94	M = 953.78, SD = 311.32	t(123) = -1.77, p = .039
Countermanding (IES)	M = 464.01, SD = 56.211	M = 481.88, SD = 65.27	t(121) = -1.63, p = .053
UFOV (threshold)	M = 50.99, SD = 15.55	M = 56.40, SD = 17.20	t(121) = -1.82, p = .035
ANT (IES)	M = 425.31, SD = 58.39	M = 474.82, SD = 63.12	t(121) = -4.52, p < .001
Mental Rotation (percent correct)	M = .80, SD = .22	M = .76, SD = .22	t(123) = 0.94, p = .174
Paper Folding (percent correct)	M = .79, SD = .25	M = .41, SD = .25	t(123) = 8.49, p < .001
RAPM (percent correct)	M = .62, SD = .18	M = .55, SD = .21	t(122) = 1.95, p = .027
UCMRT (percent correct)	M = .83, SD = .76	M = .76, SD = .20	t(123) = 2.18, p = .016
Mill-Hill (percent correct)	M = .80, SD = .15	M = .57, SD = .19	t(123) = 8.00, p < .001
Shipley (percent correct)	M = .58, SD = .12	M = .47, SD = .15	t(123) = 4.65, p < .001

Table S3
Observed and expected counts of researcher guesses of participant expectations and training conditions.

	Participant Expectation		Training Condition	
	Observed	Expected	Observed	Expected
Incorrect Guess	152	156.7	176	156.7
Correct Guess	83	78.3	59	78.3

Table S3
Moderator analyses of post-test performance on the O-span, countermanding, mental rotation, paper folding, UFOV, and ANT tasks.

	O-span			Countermanding			Mental rotation			Paper folding		
Moderator	b	p	CI	b	p	CI	b	p	CI	b	p	CI
Age	.09	.078	-.01, .20	.04	.502	-.08, .16	.01	.792	-.08, .11	.00	.969	-.11, .11
Gender	.13	.683	-.51, .77	-.00	.990	-.71, .70	.29	.319	-.28, .87	-.23	.495	-.90, .44
Big 5												
Extra.	.06	.440	-.09, .20	.01	.916	-.15, .17	-.01	.952	-.13, .12	.03	.702	-.12, .18
Agree.	.06	.450	-.12, .24	-.02	.810	-.23, .18	-.06	.468	-.23, .11	.07	.476	-.12, .26
Consci.	-.02	.852	-.20, .16	.02	.834	-.18, .22	-.10	.237	-.26, .07	.04	.647	-.15, .24
Neurot.	-.04	.592	-.20, .11	.09	.302	-.08, .26	.03	.631	-.11, .17	-.11	.200	-.28, .06
Open.	-.04	.662	-.19, .12	.03	.759	-.15, .20	-.09	.200	-.24, .05	.01	.870	-.15, .18
BAS/BIS												
Drive	.06	.328	-.06, .17	.07	.307	-.06, .19	-.03	.552	-.14, .07	-.02	.789	-.14, .11
Fun-seek	-.01	.879	-.14, .12	.02	.832	-.13, .17	-.06	.338	-.18, .06	-.04	.625	-.18, .11
Reward	.09	.152	-.03, .22	-.05	.456	-.19, .09	-.11	.067	-.22, .08	-.04	.524	-.18, .09
BIS	.01	.796	-.07, 09	-.01	.772	-.10, .07	.02	.571	-.05, .09	-.06	.160	-.14, .02
Grit	.02	.569	-.04, .07	.04	.250	-.03, .10	.03	.265	-.08, .02	.04	.225	-.02, .10
Metacog.	.01	.379	-.10, .02	.00	.656	-.01, .02	**-.02**	**.007**	**-.03, -.01**	.00	.686	-.01, .02
Mindset	**-.05**	**.013**	**-.09, -.01**	.00	.978	-.04, .04	*-.03*	*.057*	*-.07, .00*	-.01	.721	-.05, .03
SSEIT												
Percept	-.01	.634	-.05, .03	.00	.915	-.05, .05	**-.05**	**.023**	**-.08, -.01**	-.03	.191	-.08, .02
Utility	**.12**	**.003**	**.04, .19**	.03	.554	-.06, .11	-.06	.090	-.13, .01	-.06	.192	-.14, .03
Self	-.01	.773	-.06, .05	-.01	.867	-.06, .05	-.04	.097	-.09, .01	.01	.713	-.05, .07
Others	-.03	.363	-.10, .04	-.02	.524	-.10, .05	-.05	.076	-.11, .01	-.01	.792	-.08, .06
WOFO												
Work	.04	.457	-.06, .13	.04	.444	-.06, .15	**-.09**	**.037**	**-.18, -.01**	.06	.264	-.04, .16
Mastery	.04	.219	-.03, .10	.02	.573	-.05, .09	.01	.605	-.04, .07	.06	.089	-.01, .12
Compete	.02	.564	-.06, .10	-.02	.736	-.11, .07	.02	.645	-.07, .09	-.01	.868	-.09, .08

	UFOV			ANT		
Moderator	b	p	CI	b	p	CI
Age	-.01	.868	-.09, .08	-.02	.765	-.14, 10
Gender	-.15	.674	-.88, .57	.05	.901	-.69, .78
Big 5						
Extra.	-.05	.575	-.21, 12	.14	.086	-.02, .30
Agree.	.18	.081	-.02, .39	-.01	.959	-.21, .20
Consci.	-.15	.165	-.35, .06	-.10	.326	-.31, .10
Neurot.	-.06	.512	-.23, .12	-.15	.096	-.32, .03
Open.	-.11	.229	-.29, .07	-.02	.800	-.21, .17
BAS/BIS						
Drive	-.01	.843	-.14, .12	.09	.164	-.04, .22
Fun-seek	-.01	.923	-.16, .15	.05	.488	-.10, .21
Reward	-.01	.874	-.15, .13	-.02	.750	-.17, .12
BIS	.04	.410	-.06, .12	**-.10**	**.018**	**-.19, -.02**
Grit	-.04	.227	-.10, .03	.01	.788	-.06, .07
Metacog.	-.01	.302	-.03, .01	.01	.422	-.01, .03
Mindset	.00	.941	-.04, .05	-.00	.862	-.05, .04
SSEIT						
Percept	-.02	.542	-.07, .03	-.02	.392	-.07, .03
Utility	.02	.743	-.08, .10	-.02	.725	-.10, .07
Self	.05	.100	-.01, .11	.01	.717	-.05, .07
Others	.02	.675	-.06, .09	-.02	.689	-.09, .06
WOFO						
Work	.06	.276	-.04, .17	-.00	.979	-.11, .11
Mastery	.00	.970	-.07, .07	-.03	.463	-.09, .04
Compete	-.02	.702	-.11, .07	-.08	.080	-.17, .01