Journal of Cognitive Enhancement
https://doi.org/10.1007/s41465-021-00207-6

ORIGINAL RESEARCH

No Evidence for Expectation Effects in Cognitive Training Tasks

Mariya Vodyanyk[1] · Aaron Cochrane[1] · Anna Corriveau[1] · Zachary Demko[1] · C. Shawn Green[1]

Received: 9 November 2020 / Accepted: 13 January 2021

Abstract

A great deal of recent empirical and theoretical work has examined whether it is possible to enhance cognitive functioning via behavioral (cognitive) training. While a growing body of research provides support for such a hypothesis, multiple critiques of the field have suggested that any positive findings in the field to date may be due to placebo effects, rather than reflecting "true" benefits of the training paradigms. Here, in a series of four experiments, we sought to purposefully induce placebo effects of this type in cognitive training-style setup. We did so in multiple outcome domains (fluid intelligence; spatial skills), employed multiple types of "training" paradigms (classic cognitive training using the N-back working memory task; the video game Tetris) and critically, combined explicit verbal instructions that participants in some groups "should" expect to improve their performance after completing their training with associative learning "evidence" that such improvements were occurring (via manipulated task designs). In no case, though, was a placebo effect observed. These results collectively provide evidence against the contention that placebo effects are a major driver of positive outcomes previously attributed to cognitive training interventions.

Keywords Cognitive training · Expectation effects · Placebo effects · Behavioral interventions

Introduction

The past several decades have witnessed a surge in research focused on the possibility that human cognitive functions can be purposefully improved via dedicated behavioral interventions. Research in this domain has been spurred both by a multitude of basic science advances (e.g., in the field's understanding of neuroplasticity and how neuroplastic processes can be manipulated; Nahum et al. 2013) as well as by increases in real-world need (e.g., a rapidly aging Western world; an increase in professions that demand enhanced cognitive function; the ever-growing importance of formal schooling as a determinant of life outcomes; Deveau et al. 2014a; Prakash et al. 2014; Rohde and Thompson 2007; Schlickum et al. 2009; Stieff and Uttal 2015; Wright et al. 2008). Consistent with the many remaining open basic science questions, as well as the enormous number of possible translational applications, the cognitive targets of interest for such interventions have included everything from fluid intelligence, to processing speed, to perceptual skills, to spatial cognition, to a host of executive functions including attentional control, working memory, cognitive flexibility, and inhibition (Ball et al. 2002; Deveau et al. 2014b; Karbach and Unger 2014; Kramer et al. 1995; Schmiedek et al. 2010; Strobach and Karbach 2016; Valdes et al. 2017). Thus far, many researchers in the domain have argued that the existing empirical data provides reason to be optimistic that some cognitive functions can be enhanced via some forms of behavioral training (Au et al. 2015; Bediou et al. 2018).

✉ C. Shawn Green
cshawn.green@wisc.edu

[1] Department of Psychology, University of Wisconsin-Madison, Madison, WI 53705, USA

Concerns Regarding Placebo Effects as a (or the?) Mechanism Driving Positive Cognitive Training Effects

The optimistic view of cognitive training has certainly not been uniformly expressed by all researchers, with challenges and critiques coming from a variety of directions (Boot et al. 2013; Shipstead et al. 2012a; Simons et al. 2016). For instance, there has been considerable debate regarding whether the central tendencies of the existing empirical literature do in fact point in a positive direction, rather than toward a null (and

Published online: 12 March 2021

these debates have in turn often exposed differences in opinion regarding, for example, what types of studies are appropriate to aggregate across in meta-analytic work; Au et al. 2015; Melby-Lervag and Hulme 2013; Shipstead et al. 2012b). Perhaps the most significant avenue of criticism of the field to date though has focused on perceived methodological shortcomings. In particular, it has been argued that typical methodologies utilized in the field leave alive the possibility that any observed positive outcomes are the result of participant expectations (e.g., placebo effects), rather than being driven by the given behavioral training paradigms per se (Boot et al. 2013).

In addressing this criticism, it is first worth noting that the number of distinct approaches that might fit under the superordinate label of "behavioral interventions for cognitive enhancement" is vast. Yet in terms of core methodological approach, most studies in this broad field do utilize a reasonably common set of practices that in turn reflect a similar set of high-level goals. It is therefore at this level of shared broad methodological approach that we will consider the field below (Green et al. 2019; Green et al. 2014).

Specifically, the most typical research design used to assess the impact of behavioral training on cognitive function is a pre-test → intervention → post-test design with two intervention arms (experimental and control). This basic experimental design shares many features of randomized controlled trials (RCTs), as are employed in various medical domains (Price et al. 2008). Yet the very nature of behavioral interventions makes some differences impossible to avoid. In particular, in the pharmaceutical domain, it is possible to create two meaningfully different arms that are nonetheless outwardly fully identical to the participants (e.g., a "real pill" and a "sugar pill"), thus blinding participants to condition (although note that even in pharmaceutical trials, researchers do not always evaluate whether the blinding was in fact successful (Hrobjartsson et al. 2007)). In the case of behavioral interventions for cognitive enhancement, it is not possible for the two arms—experimental and control—to be outwardly identical. Indeed, if two behavioral paradigms were created that were outwardly identical, including what stimuli are presented to the participants, how the participants are asked to respond to the stimuli, etc., they would not just be "outwardly identical;" instead, they would in fact be literally identical. Simply put, if two behavioral paradigms are to differ in their impact on cognitive functioning, they must necessarily also differ in terms of their outward appearance (Green et al. 2019).

Given this inherent constraint associated with behavioral interventions, a typical approach used by investigators in the field is to attempt to create control training paradigms that participants believe are actually experimental training paradigms. In other words, because the outward appearance of the experimental and control training paradigms cannot be matched in the case of behavioral interventions, investigators instead attempt to create a situation in which the participants' expectations regarding the likely impact of the two paradigms are matched. For example, investigators in some studies have utilized an N-back working memory task as the experimental intervention and an explicit long-term memory task (e.g., learning trivia) as the control intervention (Jones et al. 2018). In other cases, investigators have used a working memory task with adaptive difficulty as the experimental intervention (i.e., where the difficulty increases as participants learn and improve) and a non-adaptive version of the same working memory task (i.e., that stays at a very easy level of difficulty for the entirety of training) as the control condition (Klingberg et al. 2005).

Yet, regardless of the specific approach that investigators have taken in the attempt to create control training paradigms that participants could view as being equally likely to impact their cognitive function as the experimental training paradigm, it will remain the case that the active treatment and the control treatment will necessarily differ in meaningful ways. This in turn leaves open the possibility that participants in cognitive training studies may be able to infer, based on the training task characteristics, whether they are in an active intervention or a placebo-control group. This possibility then serves as the foundation for a major class of criticisms leveled against the field, namely that the typical methods are inadequate with respect to participant blinding (i.e., that participants could determine whether they are in the experimental group or the control group based upon the characteristics of the training they receive). The argument then progresses to state that if participants are able to infer their group assignment, then this, combined with accurate expectations regarding how their assigned group, "should perform" (e.g., that the experimental group "should improve" or that that control group "should not improve") and the ability to purposefully alter behavior at post-test in such a way to conform with those expectations could be the true mechanism underlying positive effects observed in the field thus far (i.e., that any observed positive effects are "placebo effect").

Existing Research on Whether Expectations are a Likely Confound in Cognitive Training Studies

The current evidence supporting the type of pathway, from expectations to shifts in behavior, in cognitive training-type contexts is, at best, mixed. Here, within this small subset of the literature, it is important to distinguish between two key methodological approaches: (1) studies that have attempted to create differences in expectations across groups via biased sampling (i.e., recruiting groups that are likely to differ in the expectations they bring to a study) and (2) studies that attempt to create differences in expectations in randomly assigned groups via explicit instructions.

As an example of the former method, Foroughi and colleagues recruited participants via two types of posters (Foroughi et al. 2016). One type of poster explicitly mentioned that the study was a cognitive training study and that cognitive training could result in enhanced intelligence ("Numerous studies have shown working memory training can increase fluid intelligence…"). The other type of poster was neutral and simply indicated that participants were needed for a study ("Looking for SONA credits? Sign up for a study today"). Upon arriving at the lab, participants, regardless of which poster they had responded to, first completed a pre-test assessment of fluid intelligence. They then completed 1 session of working memory training, using the common dual N-back paradigm. Then, on the next day, the participants completed a post-test assessment of fluid intelligence. No changes in fluid intelligence scores were noted in the group recruited via the neutral expectations poster. However, large and significant improvement in scores on the fluid intelligence measures were observed in the group recruited via the expectation-inducing posters. This result, though, has not been consistently observed. For instance, utilizing a very similar biased sampling procedure, Katz and colleagues saw no such effects (Katz et al. 2018). In their case, participants were either recruited via a poster that indicated they would be participating in a cognitive training study that would improve their intelligence or via a poster that simply indicated that they would be taking part in a long-term study and would be monetarily compensated for their time. Here, the authors found that although both groups showed improvements in the fluid intelligence measure from pre-test to post-test (that was greater than a control-trained group), the two recruitment groups did not differ from one another.

With respect to the second type of approach, in which investigators attempt to deliberately create differences in expectations in randomly assigned groups, again, there are only a few example studies. In Tsai et al., participants were recruited from a common population and then were randomly assigned to receive one of two possible narrations describing the upcoming study (Tsai et al. 2018). One narration indicated that the training the participants would be receiving would result in significant improvements not just on their trained task, but on other tasks as well. The other narration indicated that the training the participants would be receiving would result in significant improvements on just their trained task, but not any other tasks. The participants then underwent 7 sessions of training on either an active (N-back) training condition or a control (trivia) training condition. The authors found that there was no difference in outcomes as a function of the initial narration condition that participants received. Similarly, Rabipour and collaborators also provided two types of information to participants prior to undergoing a training experiment (Rabipour et al. 2019). One group was told they would receive cognitive training that would globally improve their performance, while the other group was told the training was unlikely to produce any benefits. Consistent with the results above, no effects were seen as a function of the different pre-training information that was provided (although the group that was provided the positive expectations did engage more with the training).

Thus, the existing research certainly does not strongly support the contention that placebo effects are a pervasive issue in cognitive training. However, there are two important caveats that must be considered. First, the literature to date is incredibly sparse. It would thus be premature to make a strong argument on the basis of such a small number of data points. Second, the "explicitly provided expectations" approach that investigators have used in the cognitive training domain in an attempt to induce placebo effects thus far is known in domains that have more thoroughly studied placebo effects to be a weaker method of inducing such effects.

Research on Placebo Effects: Lessons from Outside Domains

Although the possibility of participant expectations playing a role in intervention outcomes has only recently gained traction in the context of cognitive training, such effects have received considerably more attention in a variety of other domains, such as pain analgesia. One key point that work in these domains has made clear is that participant expectations can be manipulated via a variety of routes (Büchel et al. 2014; Colloca et al. 2013; Kaptchuk and Miller 2015). One of these routes is via explicit instruction, as described above, where various training conditions were explicitly described to participants as being likely to improve their cognitive functions. Another route is via observational learning. Humans are excellent at learning vicariously (Bandura et al. 1966). If we witness another person receive a shot and react with pain, we will naturally tend to expect that the shot will cause us pain as well. The final route is via personal experience. In the case of pain for instance, if we find that after taking a certain pill, our headache pain is diminished, we will expect that taking the pill in the future will also lead to a reduction in pain.

Interestingly, one common tactic employing this latter "personal experience" route has involved creating false associative pairings between participants' experiences and perceived outcomes. One particularly good example of this general approach came from the study of pain analgesia (Voudouris et al. 1985). The full study in question involved three sessions. In the first session, participants were exposed to a certain painful stimulus, both with and without the application of a cream that was described to participants as an anesthetic, but was in fact, just cold cream (i.e., the cream actually had no analgesic/anesthetic properties). In the second session, participants were again exposed to a painful stimulus with and without the application of the cream. However, unbeknownst to the participants, in one group of participants, when the cream was applied, the actual level of the painful stimulus that was given was physically reduced relative to what was given during session 1. In another group of participants, when the cream was applied, the actual level of the painful stimulus that was given was increased relative to what was given during session 1. This manipulation was meant to provide personally experienced "evidence," either of the cream's effectiveness or lack of effectiveness. In essence, the purpose of the manipulation was for the cream to be associatively paired with either less pain than previously expected or more pain than previously expected. Finally, a third session progressed in a manner identical to the first. The researchers found that the associative conditioning process not only produced a strong placebo effect in general, but that it actually overwhelmed the verbally provided information when there was a conflict between these sources of information.

Overall Research Goals and Strategy

Given the existing state of the literature, there are several questions that the current research was designed to speak to. First, all studies to date in the cognitive training field have attempted to induce placebo effects purely via explicit instructions and/or recruiting materials. Given that associative learning–based manipulations appear to produce stronger placebo effects in the pain placebo literature, including such a manipulation was important given the goal of providing the best possible chance to observe a placebo effect in the context of cognitive training. Second, the little work that exists on the impact of expectations in cognitive training-type contexts has focused primarily on working memory/fluid intelligence as constructs of interest. However, within the field of cognitive training, many other constructs have been viewed as potential targets of interventions. Critically, these other targets may be more or less susceptible to expectation effects as compared to working memory/fluid intelligence. As such, we chose to explore the potential for expectations to drive changes in outcomes not only in fluid intelligence, but in spatial cognitive skills—in particular, mental rotation, which has been linked to performance in science, technology, math, and engineering fields and is thus another common target of cognitive training interventions (Uttal et al. 2013a, 2013b). Finally, within those studies that have utilized explicitly provided expectations, they have uniformly done so at the level of the task (e.g., "Cognitive training of the type that you will undergo here will…"). However, it may be important for participants to not only be told that "cognitive training has been shown to...", but for it to be further indicated that they are themselves personally capable of showing such results. Indeed, the importance of individuals believing that they themselves are capable of growth, not just that "people" are capable of growth, has been seen as a critical factor in the field focused on "growth mindset" (Dweck 2006). We, therefore in two studies (one in-person and one online) provided additional expectations at the level of the individual, in an attempt to determine if this produced a differential outcome.

In short, our goal with the following set of studies was to provide a strong possible set of opportunities to observe placebo effects in cognitive training contexts. If placebo effects are observed, it would suggest that these may be of concern in interpreting previous work in the domain (with the obvious caveat that no previous work focused on cognitive training has so clearly attempted to deliberately create placebo effects). Conversely, if no placebo effects are observed, despite utilizing methods that are far more likely to induce expectations than those employed in any existing cognitive training studies, it would provide evidence against the contention that such effects are a likely major confound in previous cognitive training work.

Study 1: Assessing Whether Positive Explicit Expectations + Positive Associative Learning Evidence Induce Placebo Effects in a Measure of Fluid Intelligence

Previous work attempting to induce placebo effects in cognitive training-like setups has done so exclusively via explicitly provided verbal suggestions. Here, we sought to examine the impact of verbal suggestions combined with an associative learning manipulation that, like the pain placebo study described above, was meant to provide personally experienced evidence of the effectiveness of the intervention. In the pain placebo study, such evidence was provided by, unbeknownst to the participant, manipulating the level of the painful stimulus that was delivered to make it seem as if the cream was reducing the level of experienced pain (i.e., telling the participant that they were being given the standard level of painful stimulus, but in fact giving them a less painful stimulus). Here, we took an analogous approach by, unbeknownst to the participants, manipulating the difficulty of matrix reasoning trials that they were provided halfway through their training. Specifically, at the training mid-point, participants were given a set of matrix reasoning trials that was described as being equivalent to the set of trials they completed at pre-test, but was in fact, normatively easier, so as to provide "evidence" of the effectiveness of the training.

Methods

Participants

For this and all subsequent studies, participants were UW-Madison undergraduate students enrolled in Introduction to Psychology. Participants were recruited via an online (SONA) system. All recruitment materials (e.g., study titles/descriptions) were made purposefully ambiguous to ensure that participants who chose to enroll in the study were unaware of the study conditions or purpose. All participants provided informed consent and received extra course credit for participating. All studies were approved by the University of Wisconsin Education and Social/Behavioral IRB. For any conditions that involved deception (e.g., suggesting to participants that they had improved, when in fact, they were provided easier trials), participants underwent a full debriefing at the conclusion of the study in accordance with the approved IRB protocol.

For study 1, 145 participants were enrolled (average age = 19; 98 females) and were randomly assigned to one of two conditions: (1) the positive verbal expectations + positive associative expectations condition (N = 74) or (2) the control condition (N = 71). For ease of exposition, under study 1, we will refer to the former simply as the "expectations" group and the latter as the "control" group (see supplemental Materials for information about sample sizes across all studies).

Procedure

The study utilized a "pre-test (matrix reasoning) → training (N-back working memory training) → mid-test (matrix reasoning) → training (N-back working memory training) → post-test (matrix reasoning)" design. Participants assigned to the expectations group were given the following script to describe the study before the pre-test. This script was specifically designed to give participants the explicit expectation that the N-back training they would be engaging in was likely to increase their performance on the fluid intelligence matrix task:

> This study looks at the effects of a specific kind of training on intelligence. The task you will be training on is called the n-back. Many published papers have shown that training on this task increases people's intelligence scores. We've designed a version of the n-back task that we think will be even more powerful. There will be two training blocks, with a short break halfway through each block. Your progress in intelligence improvement will be tested through blocks of matrix tests in-between and after these training blocks. There will also be an initial matrix test before you start your training as a baseline measure. If the training is working, you should find yourself doing better and better on these matrix tests as the training goes on. Each matrix test will have 18 questions and you will be given 20 minutes to complete them.

After reading the above information, participants then completed a brief mindset questionnaire (as a possible individual difference factor; note that purposefully manipulating perceptions of growth mindset is the focus of study 4; see also Supplemental Tables S1 & S2). Participants then completed a pre-test set of 18 matrix reasoning problems (Pahor et al. 2018). These matrix reasoning problems have previously been normed to have varying levels of difficulty, and the specific matrices chosen for the pre-test set were selected to provide a reasonably wide range of difficulties (expected number of items correct was approximately 10 out of 18 given the previous data). After each trial of the matrix test, participants were given feedback regarding both whether they were correct and, if not, which option was correct. After either 20 min or after the participant completed all 18 problems, they were redirected to a page that showed them the total number of questions they got correct out of 18.

Following the completion of the pre-test, participants then moved on to the first round of training on the N-back task. The version of the N-back task that was employed was a visuospatial version, where participants viewed sequences of squares that appeared pseudorandomly within the spatial positions defined by a 3 × 3 grid (other than the center position). Their task, throughout the entirety of training, was to indicate whether the position of the current square that was presented was the same as that of the square 3-back in the sequence (half trials = yes, half trials = no). Participants completed 140 trials total in the first bout of training (i.e., where each response = a trial), with a short break occurring halfway through. In all, this first bout of training lasted approximately 5–8 min.

Participants then completed another 18 matrix reasoning trials as a mid-test. This proceeded in a manner that was identical to the pre-test with the exception of the difficulty of the problems. The matrix problems used for the mid-test were specifically selected to be easier on average than the 18 trials that were taken at pre-test (expected percent correct for the 18 mid-test items was 70%).

The participants then completed another 140 trials of the N-back task, again with a break halfway through, before completing the post-test matrix reasoning trials (18 trials; expected level of difficulty approximately matched to the pre-test; 52%). Participants were then asked a few questions about their study experience (e.g., with regard to whether they had suspicions regarding the true study purpose), were debriefed regarding the deception, and were specifically asked to not discuss the study with any other individuals (to avoid tainting possible future participants).

Participants in the control condition underwent a reasonably similar overall procedure. The main differences were the following: (A) they were told the purpose of the study was simply to examine whether performance on the two tasks was related (i.e., at no point was the N-back referred to as "training"—it was simply treated like a task, in the same vein as the matrix reasoning task) and (B) the control participants did not complete a mid-test. Instead they took a 15-min break in between the first and second 140-trial blocks of the N-back task. This was done for two reasons: (1) to minimize the possibility that participants viewed one task as training and the other as repeated testing and (2) any matrix reasoning tasks given during the mid-test could be viewed as providing evidence of either positive change (if the same set of easier trials as was given to the expectations group was used) or lack of change (if a set of trials matched to the pre-test was used). We note that this design choice could potentially have biased results in favor of the expectations group by, in essence, giving the expectations group more practice with the matrix reasoning task. As such, if significantly larger pre-test → post-test improvements in matrix reasoning scores were observed in the expectations group as compared to the control group, this would have necessitated an additional control group that also performed the mid-test. However, as we will see in the results, this was not the case and therefore no such follow-up control was needed.

Results

We first examined the impact of our expectation induction manipulations. Significantly more participants in the expectations group (47%) as compared to the control group (19%) reported believing that the n-back training caused them to improve on the matrix tests when probed during the debriefing period ($X^2 = 12.81$, $p < 0.001$). The associative learning manipulation was similarly successful (i.e., that performance was higher in the mid-test than in pre-test/post-test). Paired samples t tests confirmed participants scored higher on the mid-test ($M = 14.73$) than either the pre-test ($M = 10.04$), $M_{\text{diff}} = 4.68$ [CI: 4.05, 5.32] or post-test ($M = 11.19$), $M_{\text{diff}} = 3.53$ [CI: 2.9, 4.17]. Although not a necessary part of the manipulation, we further found that participants were largely explicitly aware that they had performed the best in the mid-test (85% of participants reported believing they did the best on the mid-test, which is significantly greater than would be expected by chance; $X^2 = 82.62$, $p < 0.001$).

We then examined our core research question. Given our experimental design, a placebo effect would manifest as better post-test scores on the matrix reasoning task at post-test in the expectations group as compared to the control group (after controlling for pre-test scores). Counter to such a prediction though, no such group difference was observed (see descriptive statistics and full model results in Table 1 and Fig. 1). We approximated each coefficient's Bayes factor (BF) in this model (and all models without random effects structures) using the R package BayesFactor (Morey et al. 2018). We report all BF on a log scale, with a logarithm base of 3, so that $BF_{\log3} > 1$ can be interpreted as "at least moderate evidence for the inclusion of the coefficient in the model" and $BF_{\log3} < -1$ can be interpreted as "at least moderate evidence for the null hypothesis." Absolute $BF_{\log3}$ over 2.1 can be interpreted as "at least strong evidence" for either the alternative or null hypotheses, depending on the sign of the $BF_{\log3}$ (Jeffreys 1961; Wetzels et al. 2011).

Discussion

Despite providing participants with the strong verbal explicit expectation that the N-back training should enhance matrix reasoning performance, as well as providing them with associative evidence of this positive relationship via a deceptively manipulated mid-test, we nonetheless observed no reliable placebo effect. Instead, we observed moderate evidence for a lack of differences between conditions. This result is consistent with those of both Tsai and colleagues as well as Rabipour and colleagues, who likewise observed no placebo effect in a cognitive training context (Rabipour et al. 2019; Tsai et al. 2018). Our results add to that previous work by showing that a placebo effect was not generated even when including a method of placebo-induction (associative learning) that in outside domains is typically even stronger than verbally provided information alone.

Study 2: Assessing Whether Positive Explicit Expectations + Positive Associative Learning Evidence Induce Placebo Effects in a Measure of Spatial Skills

In study 1, no evidence of a placebo effect was found in a setup employing N-back training and matrix reasoning pre-tests/post-tests. Yet, while working memory training is one training type commonly employed in the realm of cognitive training, and fluid intelligence measures such as matrix reasoning tasks are a common dependent variable, these are far from the only training types or dependent variables in the field. Indeed, although in study 1 we observed negligible evidence that matrix reasoning tasks are susceptible to placebo effects, it may be possible that other cognitive domains are more impacted by participant expectations. Furthermore, it is possible that participants may have stronger a priori beliefs about the possible impact of other forms of training as compared to N-back training. Thus, in study 2, we sought to explore both a new type of training (video games) and a new dependent variable (spatial skills/mental rotation). Video games are an interesting testbed as results related to the impact of video game playing on a variety of human behaviors are often featured in popular news sources, which might promote larger expectations than N-back training (Bavelier et al. 2018;

Table 1 Study 1 outcomes (A. Participant Performance; B. Model Results)

A. Descriptive statistics			
	Pre-test	Mid-test	Post-test
Control group	10.6 [9.8,11.5]	NA	11.2 [10.4,12.1]
Expectations group	10.0 [9.3,10.8]	14.7 [14.0,15.3]	11.2 [10.3,12]
B. Model results			
	Estimate	*t* value	*p* value [BF_{log3}]
(Intercept)	3.058	3.582	0.000
Pre-test	0.767	10.556	0.000 [33.24]
Group (expectations/control)	0.435	0.912	0.364 [− 1.72]

Bediou et al. 2018). Spatial skills are of note as a possible target of cognitive training as they are often more predictive of academic performance (in particular in STEM fields) than other cognitive processes (Uttal et al. 2013a, 2013b). We thus conducted a study analogous to study 1 above, but utilized a mental rotation task as the key dependent variable, and a video game (Tetris) with, on the surface, clear links to mental rotation as the experimental training task.

Methods

Participants

For study 2, 74 participants were enrolled (average age = 19; 51 females) and were randomly assigned to one of two conditions: (1) the positive expectations group ($N = 37$) or (2) the control group (neutral/unblinded) ($N = 37$). As in study 1, under study 2, for ease of exposition, we will refer to group 1 simply as the "expectations" group and group 2 as the "control" group.

Procedures

The basic research design was analogous to that described above: pre-test → training → mid-test → training → post-test. The major changes from experiment 1 were with respect to the test (mental rotation) and training (Tetris).

The task used for testing the expectations manipulations in study 2 was the mental rotation task. On each trial, two complex, 2-dimensional shapes were presented. The objects were either identical copies, but with one rotated relative to the other, or they were mirror-reversed and rotated copies (Fig. 2). Participants were asked to report whether the objects were identical or mirror-reversed. Performance on mental rotation tasks is known to be a clear function of the difference in angle between the objects. Specifically, participants are less accurate and slower to make decisions when the objects are rotated by a larger degree relative to one another than if the objects are rotated by a lesser degree relative to one another. During the pre-test, rotation angles were chosen from a set of 9 constants spanning the range from 0 to 180° (10, 30, 50, 70, 90, 110, 130, 150, and 170°). Participants received feedback after each response for accuracy and reaction time, with the feedback presented on screen for 1 s.

Each block included 252 trials of mental rotation. These trials included 7 trials each of every combination of (each of the 9) orientations, clockwise/counterclockwise rotation, and mirror-reversed/not, presented in randomized order.

After the mental rotation pre-test, participants then completed 5 min of the video game Tetris. Participants in the expectations group were explicitly told that Tetris had been shown to improve mental rotation performance and thus they would be expected to improve at the mental rotation task after this training. The script follows:

> Previous studies have shown that practice on games that involve mentally rotating shapes improves performance on more complex cognitive rotation tasks. After

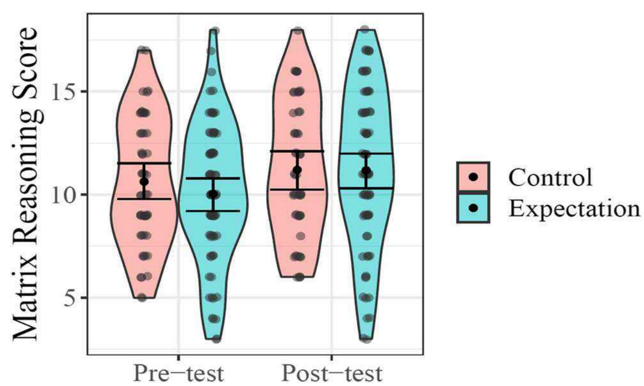

Fig. 1 Study 1 results—no evidence of placebo effect from combined verbal and associative learning expectation induction procedures: matrix reasoning scores (out of a maximum of 18) are plotted as a function of group (expectations/control) and test (pre-test/post-test). On this plot, a placebo effect would be indicated by a larger increase from pre-test to post-test in the expectations group than in the control group. However, this pattern of results was not observed. While participants showed a numerical increase from pre-test to post-test, this effect was not significantly different in the two groups. *Error bars indicate bootstrapped 95% CI of the mean

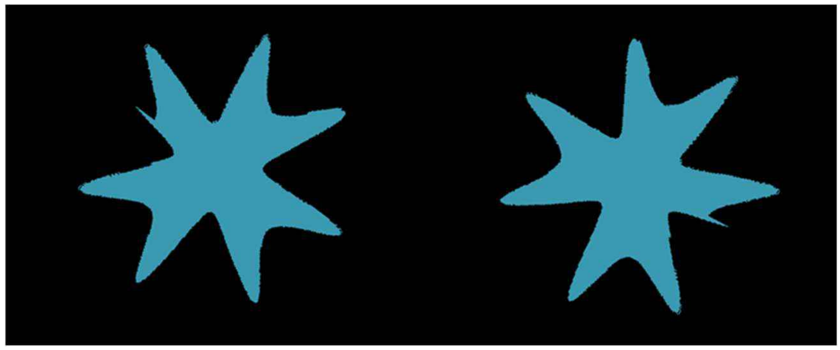

Fig. 2 Mental rotation task: on each trial, participants observed a pair of complex shapes. The shapes were either rotated copies of one another or else were mirror-reversed and rotated copies of one another. The participants' task was to indicate whether the shapes were the same or mirror-reversed copies

> 5 minutes of playing Tetris, we will repeat the mental rotation task a second time in order to determine whether the game has improved performance.

Participants in the control group, meanwhile, were explicitly told that they were in a control group and that while other participants were being erroneously told that Tetris was likely to enhance their mental rotation performance this was not in fact true:

> Previous studies have shown that practice on games that involve mentally rotating shapes improves performance on more complex cognitive rotation tasks. However, 5 min will not be long enough to improve your ability on our mental rotation task. You are a control for a group that was led to believe their performance would improve from the small amount of practice with Tetris. However, please still perform to the best of your abilities to ensure meaningful data is collected.

After this initial bout of Tetris training, participants repeated the mental rotation task. As in study 1 above, the mid-test was purposefully made easier than the pre-test. Here, this was accomplished by reducing the range of orientation offsets between the objects (0–90°). Instead of presenting one of 9 angles with a uniform probability, in the mid-test block, 44% of trials were rotated by 10°, 44% were rotated by 30°, and the remaining 11% were rotated by 90°. This manipulation was expected to produce significantly better (faster/more accurate performance) without it being obvious to participants why their performance was increased.

After the mid-test, participants played another 5 min of Tetris before completing the post-test mental rotation task (same difficulty as the pre-test).

Following the post-test, participants were probed about their perceived performance by being asked to rank the three blocks in terms of best to worst performance and by asking whether they felt any positive impact of the Tetris training. Additionally, participants were asked whether they thought we might be testing hypotheses other than what had been initially explained, and whether they had discussed their participation in the study with any student who may have participated previously. At the conclusion of the study visit, all participants were given a debriefing form which explained the true manipulation of the experiment.

Results

As in study 1, we first checked our expectation manipulations. When asked whether they thought we might be testing any alternative hypotheses, no participants in the expectations condition correctly guessed that we were examining the placebo effect. Further, significantly more participants in the expectations group (78%) than the control group (34%) reported believing that playing Tetris caused them to improve on the mental rotation tests ($X^2 = 12.44$, $p < 0.001$). The associative learning manipulation was again similarly successful (71 out of 74 participants had better performance on the mid-test than the pre-test). As in study 1, participants largely appeared to be explicitly aware of this fact. Combining both conditions, 84% of participants reported believing they did the best on the mid-test, far higher than would be expected if participants named the test that they performed the best on at random ($X^2 = 75.03$, $p < 0.001$).

We then examined our core research question. Because the mental rotation task involved asking participants to respond as quickly and accurately as possible, we chose to utilize a dependent variable that took into account both speed and accuracy. Specifically, response times/choices were modelled as arising via a drift diffusion process and thus fit to a drift diffusion model (DDM; note that the results are qualitatively similar when analyzing either RT or accuracy independently, see Supplemental Tables S3 and S4). Specifically, using the R package Rwiener (Wabersich and Vandekerckhove 2014), we fit a 4-parameter Wiener DDM to each participant's overall data. Then, holding 3 parameters constant at their overall level (non-decision time, boundary separation, and bias), we fit a drift rate to each rotation angle within each block. We then used a linear mixed effects model to test condition differences in post-test drift rates while controlling for pre-test drift rates and participant-level intercepts. We approximated each coefficient's BF_{log3} using BIC comparisons (Wagenmakers 2007).

Thus, in this analysis, a placebo effect would manifest as a faster drift rate (corresponding faster RT and/or more accurate choices) at post-test in the expectations group, as compared to the control group (after controlling for pre-test scores). Counter to such a prediction, but consistent with the results of study 1, no such group difference was observed (see descriptives and full model results in Table 2 below; also see Fig. 3). In fact, the effect appears to go in the opposite direction of a placebo expectation effect (i.e., better post-test performance in the control group). However, because we had no a priori expectation of such a directional effect, we choose here to treat that outcome as equivalent to a null.

Discussion

The results of study 2 were consistent with those observed in study 1. There was no evidence that the participant expectation condition affected the measured outcome in the direction that would indicate a placebo effect. In practice, individuals in the control group demonstrated a larger increase from pre-test to post-test than did the individuals in the expectations group.

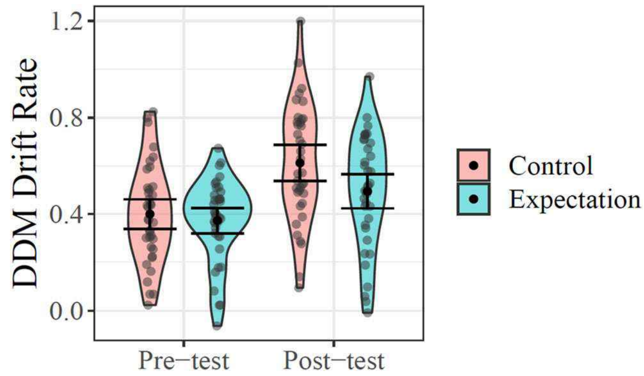

Fig. 3 Study 2 results—no evidence of placebo effect from combined verbal and associative learning expectation induction procedures on mental rotation performance: mental rotation DDM drift rate (i.e., combining speed and accuracy; higher scores = faster/more accurate responses) is plotted against test for each of the two groups. As in study 1, no evidence of a placebo effect was observed. If anything, the control group showed a greater change from pre-test to post-test than the expectations group. However, as we had no prediction that this pattern would occur, we choose to interpret this as equivalent to a null result. *Error bars indicate bootstrapped 95% CI of the mean

Study 3: Assessing Whether Positive Explicit Expectations + Positive Associative Learning Evidence Induce Placebo Effects in a Measure of Spatial Skills-"Drug" Intervention

The results of study 1 and study 2 provided no support for the contention that manipulations meant to alter participant expectations in turn serve to alter performance in cognitive training inspired contexts. Study 1 provided moderate evidence for the null result (i.e., no difference between conditions), while study 2 actually provided evidence for the opposite effect (i.e., greater improvements in the control group). As a final manipulation, we examined whether a non-behavioral induction method could induce shifts in performance. In particular, we surmised that participants perhaps had more exposure to the possibility that alterations in behavioral capacities could occur in response to substances (e.g., drugs) than from behavioral training. Indeed, there is existing research showing that the US population is increasingly aware of pharmacological interventions (e.g., through direct advertising) and this may in turn be increasing the magnitude of beliefs about the efficacy of drugs (Tuttle et al. 2015). We thus conducted an identical study to that described in study 2, but rather than utilizing

Table 2 Study 2 outcomes (A. Participant Performance; B. Model Results)

A. Descriptive statistics			
	Pre-test		Post-test
Control group	0.39 [0.33,0.47]		0.61 [0.54,0.69]
Expectations group	0.37 [0.32,0.43]		0.42 [0.32,0.49]
B. Model results			
	Estimate	*t* value	*p* value [BF_{log3}]
(Intercept)	0.130	5.20	< 0.001
Pre-test	1.204	27.83	< 0.001 [138.99]
Group (expectations/control)	− 0.085	− 3.33	0.001 [1.80]

Note that the pattern of results is qualitatively the same if just accuracy or just reaction time is analyzed instead of drift rate (see Table S3 & S4).

Note the difference in the number of orientations makes comparisons of drift rate between mid-test and pre-test/post-test problematic. However, raw RTs and accuracies for all three tests can be seen in Tables S3 & S4.

Degrees of freedom (used in calculating *p* values) were approximated using the Kenward-Rogers method.

Square brackets indicate bootstrapped 95% confidence intervals of participant means (in descriptive statistics) and BF_{log3} (in model results) behavioral training, participants were instead told that they would be ingesting a "neuroenhancing fluid" (which was strongly implied to be a stimulant, but was, in fact, only water).

Methods

Participants

For Study 3, 30 participants were enrolled (average age = 19; 15 females) and were randomly assigned to one of two conditions: (1) the positive expectations group ($N = 15$) or (2) the control group (neutral/unblinded) ($N = 15$). One participant in the positive expectations group declined to continue the study after it was explained that the study involved ingesting an unknown substance, and is therefore not included in the analysis. Again, for ease of exposition, we will refer to group 1 simply as the "expectations" group and group 2 as the "control" group.

Procedures

The basic research design was analogous to that described in study 2. As in study 2, the study began by participants completing the mental rotation pre-test. The only difference in the study 3 procedures was that no behavioral training was provided. Instead, at the two points in previous studies when participants were asked to complete a form of behavioral training, in study 3, they were asked to drink a small amount of liquid. This liquid was described as a "neuroenhancing fluid" to those participants in the expectations group. Specifically, participants were told to expect that drinking this liquid would speed their reactions and improve their mental rotation ability. They were also told that they may experience other minor side effects (e.g., increased heart rate, flushing, slight trembling) consistent with a stimulant. After this was explained to the participants, they watched as 5 ml of filtered water was measured into a graduated cylinder and then poured into a paper cup. The "neuroenhancing fluid" was then taken from a brown glass laboratory bottle with a fake chemical label for plausibility. Two milliliters of this liquid were then pipetted into the paper cup within view of the participants. Participants in the control condition meanwhile were told that the liquid was just water and that they were the control for a group that was being told the liquid was a neuroenhancing fluid. After participants drank the contents of the paper cup, they waited for a period of 5 min. The participants in the expectations group were told the neuroenhancer would take this long to metabolize.

To add additional "evidence" of the effectiveness of the neuroenhancer, a secondary test (simple reaction time) was also employed in study 3. On each trial, the participants were shown an arrow facing left or right and their task was simply to press the corresponding arrow key as quickly as possible. Participants received both accuracy and reaction time feedback after each trial. However, during the pre-test, the reaction times indicated in this feedback were manipulated by adding 50 ms to the true reaction times. This change was meant to maximize feelings of improvement during the mid-test.

After 5 min had passed, participants performed the reaction time task a second time. This time, however, 50 ms were subtracted from their times with the goal of providing "evidence" of improved performance/that the "neuroenhancing fluid" had taken effect. The experimenter verbally noted that the participant's reaction time had improved and that he/she would be allowed to move on to the second block of the mental rotation task (manipulated in the same manner as the mid-test described in study 2). After participants completed the mental rotation mid-test, they then ingested another round of the "neuroenhancing fluid," waited 5 min, and took the reaction time task/mental rotation task a final time.

Following completion of all three blocks, experimenters asked participants to rank the blocks from their best to worst performance. Participants were asked whether they felt the effects of the "neuroenhancer" on their performance on the tasks. They were probed for suspicion of a placebo study by asking if they could think of other hypotheses that may have been tested, and whether they had discussed their participation in the study with anyone who may have participated previously. Finally, all participants were given a debriefing form, which explained the true intended manipulation of the experiment.

Results

As in studies 1 and 2, we first examined our expectation manipulations. A total of 80% of the participants in the expectations group said they believed that consumption of the fluid improved their performance on the mental rotation task. However, when asked whether they thought we might be testing any alternative hypotheses, six participants correctly guessed that we were examining the placebo effect. Interestingly, five of these six were also individuals who had said they believed that consumption of the fluid improved their performance. This apparent mismatch in expressed beliefs highlights the extreme difficulty in assessing subjective feelings of improvement in behavioral training participants (i.e., that they can give mutually exclusive responses). As in studies 1 and 2, the associative learning manipulation was successful (all participants' mean response times on the mental rotation mid-test were lower than their response times on the pre-test). Again, this appeared to be explicitly recognized by the participants. Combining both conditions, 100% of participants reported believing they did the best on the mid-test, far higher than would be expected if participants named the test that they performed the best on at random ($X^2 = 60$, $p < 0.001$). We then examined our core research question using equivalent analyses as in study 2. As in study 2, no group difference consistent with a placebo effect was observed (see descriptives and full model results in Table 3 below; also see Fig. 4; note that the results are qualitatively similar when analyzing either RT or accuracy independently, see Supplemental Tables S5 and S6).

Discussion

As in the first two studies, no evidence of a placebo effect was observed when using a "drug" placebo-type manipulation. We observed moderate evidence for the null effect (i.e., no difference between conditions).

Study 4: Attempting to Produce a Placebo Effect via Manipulation of Growth Mindset-In Person and Online

While the results of studies 1–3 failed to provide support for previous suggestions that participant expectations act as a substantial confound in cognitive training studies, as noted in the introduction, one could argue that the critical ingredient is not that participants are led to believe that "some people" (or even "most people") benefit from the training. Instead, it could be critical that they believe that they themselves are capable of benefiting from the training. Indeed, the literature on growth mindset makes clear that beliefs relative to oneself, rather than relative to the population at large, are critical in predicting individuals' future behavior. This idea could also be viewed as potentially consistent with the results of Foroughi and colleagues, as one might expect that individuals who believe they are capable of benefiting from training would be more likely to respond to a poster advertising such training (Foroughi et al. 2016).

Growth mindset was measured in study 1. However, no trend toward a relation between growth mindset and magnitude of change from pre-test to post-test was observed (see Table S2). Yet, we felt a potentially stronger test of the hypothesis would be to attempt to directly manipulate participants' beliefs about their capacity for change in a positive or negative manner. Thus, in study 4, we combined the previously employed positive verbal expectations regarding the task as well as the positive associative learning with a third manipulation. Namely, participants were randomly assigned to either receive a script indicating that they were an individual who was extremely capable of behavioral change via training or a script indicating that they were less susceptible to change via experience (both framed in a positive manner, see below). Furthermore, we ran separate cohorts of participants in-person and exclusively online. While some differences between in-person and online studies might lead to the prediction of smaller placebo effects online (e.g., reductions in effects that arise via in-person observation-obedience effects; Hawthorne effect), others might lead to the prediction of larger

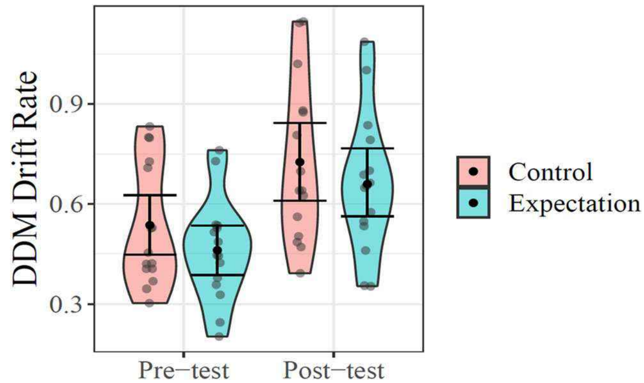

Fig. 4 Study 3 results—no evidence of placebo effect from combined verbal and associative learning expectation induction procedures on mental rotation performance. As in study 2, while participants improved their performance on the mental rotation task from pre-test to post-test, the magnitude of this effect was equivalent in the control and expectations groups. *Error bars indicate bootstrapped 95% CI of the mean

Table 3 Study 3 outcomes (A. Participant Performance; B. Model Results)

A. Descriptive statistics			
	Pre-test		Post-test
Control group	0.54 [0.44,0.63]		0.73 [0.61,0.84]
Expectations group	0.46 [0.39,0.54]		0.66 [0.56,0.77]
B. Model results			
	Estimate	*t* value	*p* value [BF_{log3}]
(Intercept)	0.009	0.22	0.824
Pre-test	1.336	24.18	< 0.001 [142.82]
Group (expectations/control)	0.034	0.84	0.406 [− 1.21]

Note that the pattern of results is qualitatively the same if just accuracy or just reaction time is analyzed instead of drift rate (see Tables S5 & S6).

Note the difference in the number of orientations makes comparisons of drift rate between mid-test and pre-test/post-test problematic. However, raw RTs and accuracies for all three tests can be seen in Tables S5 & S6 effects online (e.g., it is possible that participants would consider instructions that come from a "psychology research lab" to have more authority than instructions that are provided by an undergraduate research assistant).

Methods

Participants

For in-person, 117 participants were enrolled (average age = 20; 68 females) and were randomly assigned to one of two conditions: (1) the positive mindset condition ($n = 59$) or (2) the negative mindset condition ($n = 58$).

For online, 93 participants were enrolled (demographics could not be recorded due to data collection issues) and were randomly assigned to one of two conditions: (1) the positive mindset condition ($n = 46$) or (2) the negative mindset condition ($n = 47$).

Procedures

The general methods were identical to those described for the "expectations" group in study 1 with one exception. In study 4, after participants took the initial mindset questionnaire, they were randomly chosen to receive one of two possible scripts that they were led to believe reflected the pattern of responses they provided on the questionnaire. The script they were given was in fact independent of how they responded to the mindset questionnaire. One script was designed to give participants the verbal expectation that they themselves change easily as a function of experience (positive induction/growth mindset). The other script was designed to give participants the belief that they themselves do not change easily as a function of experience (negative induction/fixed mindset). Critically though, both were framed to the participants as reflecting desirable traits (i.e., as a "good" trait to have).

> Script #1: Positive induction (growth mindset): Your quiz results show that you have a strong growth mindset, which means that you are able to easily and seamlessly change in response to minor changes in your environment. This trait is particularly valuable in fast paced, high stress situations where you are able to adapt and grow your skills quickly to meet new challenges. The theory of growth mindset was developed by well-known psychologist Carol Dweck. Her research has shown that individuals with strong growth mindsets tend to reach higher levels of life achievement than individuals whose personality is more fixed. These higher levels of life success are largely attributed to the fact that high growth mindset individuals tend to take on new encounters, even difficult ones, with positivity, which allows them to rise to meet all occasions.

> Script #2: Negative induction (fixed mindset): Your quiz results show that you are thoughtful, steady, and not prone to be overly influenced by minor or temporary changes in your environment. This means you're less likely to be influenced by short psychology-lab type experiments such as this one. This fixed trait is uncommon among individuals and is seen as valuable in fast paced, high stress situations. For instance, multiple research studies have now shown a positive correlation between this more fixed and unchanging personality type and high performance under stress. Similarly, a recent study showed that the most successful and liked CEOs of Fortune 500 companies scored highest in this personality trait, with these individuals' employees disproportionately mentioning stability and reliability as their boss's most important traits.

Results

The key question in study 4 was whether matrix reasoning performance was significantly higher at post-test in the positive mindset group as compared to the negative mindset group, after controlling for pre-test scores. We first examined whether it was justified to combine across the in-person and online versions. There was not a significant difference between in-person and online participants in their pre-test scores (mean in-person = 10.5, mean online = 10.8, CIdiff = [−1.23,0.59], $t = -0.7$, $p = 0.5$); however, there was a significant difference in terms of their change from pre-test to post-test with the in-person showing greater improvements from pre-test to post-test than the online participants (gain score in-person = 1.7, gain score online = − 0.34, CIdiff = [2.82,1.26], $t = 5$, $p < 0.001$). Although these effects did not interact with mindset manipulation in both groups combined, given the difference in change scores we chose to examine the groups separately. In neither case, though, was there a significant impact of mindset manipulation (in-person: $b = 0.42$, $t = 0.86$, $p = 0.391$; online: ($b = -0.21$, $t = -0.34$, $p = 0.733$) (Fig. 5 Table 4).

Discussion

The results of Study 4 were consistent with those observed in study 1. In study 1, we saw no relationship between measured growth mindset and change in performance from pre-test to post-test. In study 4, there was no observed impact of attempting to manipulate this belief—in other words, providing the verbal suggestion to participants that they might be particularly capable of changing in response to their upcoming training versus providing the verbal suggestion that they might be particularly incapable of changing in response to the upcoming training.

Table 4 Study 4 outcomes (A. Participant Performance; B. Model Results)

A. Descriptive statistics			
	Pre-test		Post-test
Negative group in-person	10.7 [9.9,11.5]		12.2 [11.4,13.2]
Negative group online	11.1 [10.2,12.0]		10.8 [9.7,10.7]
Positive group in-person	10.9 [10.0,11.8]		12.8 [11.8,13.8]
Positive group online	9.8 [8.9,10.7]		9.4 [8.4,10.7]
B. Model results			
	Estimate	*t* value	*p* value [BF_{log3}]
In-person: (intercept)	4.437	5.16	< 0.001
In-person: pre-test	0.729	9.95	< 0.001 [29.50]
In-person: group (pos-neg)	0.421	0.86	0.391 [− 1.72]
Online: (intercept)	1.135	0.97	0.337
Online: pre-test	0.869	8.83	< 0.001 [22.39]
Online: group (pos-neg)	− 0.212	− 0.34	0.733 [− 1.87]

Conclusions and Future Directions

In a series of 4 experiments, together involving over 400 participants, we found no support for the contention that expectations are a likely significant confound in cognitive training designs. We assessed the potential for expectations to alter performance in cognitive-training style setups in multiple cognitive domains (fluid intelligence, mental rotation), utilizing multiple types of "training" paradigms (N-back task, Tetris) and critically employing both explicit verbal instruction and associative conditioning in an attempt to maximize the chances that a placebo effect would be observed. Yet, across all four studies, the only reliable outcome was one in the opposite direction of that predicted by the researchers who have posited that placebo effects play a major role in cognitive training outcomes. Bayes factors indicated moderate evidence for the null hypothesis (i.e., no difference between expectation and control groups) in both training paradigms. We also found no support for the idea that individuals' growth mindset impacted the extent to which a placebo effect was observed—either when growth mindset was measured or when it was manipulated.

One clear limitation of this work is the short "training" period utilized in all studies (i.e., single session). While this same approximate length has been used in prior work on expectations in cognitive training (Foroughi et al. 2016), it is far shorter than is used in any "actual" cognitive training. While our participants demonstrated consistent belief in our expectation induction materials, it may be the case that this effect is stronger in longer-term training studies or that individuals who are willing to participate in long-term training studies may be different in some ways compared to the participants utilized here. A second limitation is that our participants were exclusively college-aged young adults. There would thus be great virtue in similar follow-up work across a wider range of demographic profiles. For instance, older adults (especially those experiencing cognitive decline) might be more personally invested in improving their cognitive function and thus show a different pattern of placebo responsiveness than our young adult sample. Finally, as there was no evidence for an overall placebo effect in our data, there is little opportunity to explore whether the various manipulations (e.g., explicitly provided expectations, novel experiences, manipulated feedback) correspond to the same or different underlying mechanisms. For instance, our group has recently proposed that these various manipulations could potentially be understood in a common global framework analogous to Bayesian cue combination (e.g., where different types of cues provide evidence in favor of various beliefs with different magnitudes and degrees of certainty; Denkinger et al. in press). Testing this framework though (e.g., by pitting various manipulations against one another) would require a paradigm

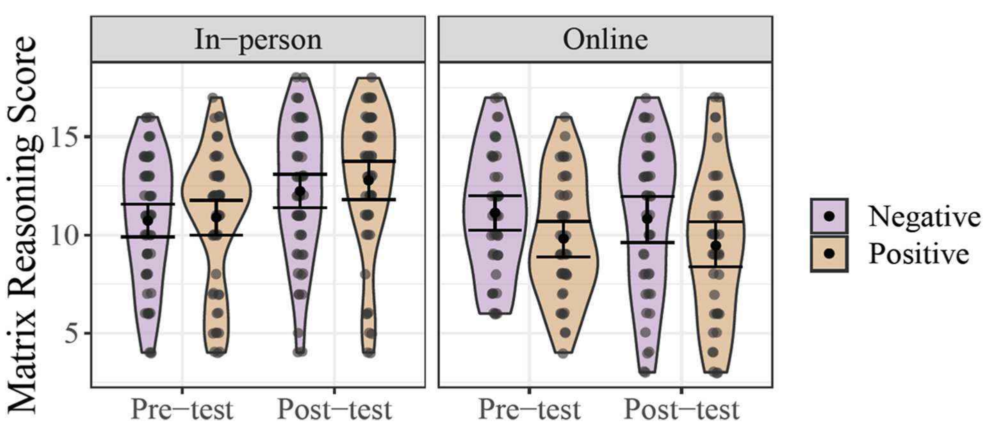

Fig 5 Study 4 results—no evidence of growth mindset manipulation on changes in matrix reasoning performance: matrix reasoning number correct is plotted against test time for each of the two groups (i.e., positive or negative growth mindset induction) and each location (i.e., online or in-person). As in study 1, no relation between growth mindset and training efficacy was observed. The divergent patterns of results between the two locations led us to consider these two sets of results independently (i.e., improvements in-person are likely to be qualitatively as well as quantitatively different than stability or decreases observed online)

that produces significant placebo effects (perhaps, as noted above, by using longer-term training or participant populations that are more susceptible to placebo effects).

Thus, in all, the current work certainly does not rule out the hypothesis that placebo effects play a role in true cognitive training designs that take place over many days (although we are not aware of any theoretical perspective that would make a distinction between expectations as a function of training duration). Yet, the work also provides no support for the idea that such effects are a "pervasive problem" in the literature.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s41465-021-00207-6.

Authors' Contributions M.V., A.Corriveau, A.Cochrane, and C.S.G. contributed to the development of the methods, procedures, and core hypotheses. M.V., A.Corriveau, and Z.D. took primary responsibility for data collection. A.Cochrane took primary responsibility for data analysis, with contributions from the remainder of the author list. All authors contributed to the writing of the final manuscript.

Funding The project was partially supported by Office of Naval Research grant N00014-17-1-2049 to C.S.G.

Data Availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request. These datasets along with all final analysis code will be uploaded to an open-data repository upon publication.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethics Approval All studies described here were approved by the University of Wisconsin-Madison Education and Social/Behavioral IRB.

Consent to Participate All participants provided informed consent to participate in the studies described here.

Consent to Publish Participants granted their consent to publish data, in the format utilized in the current manuscript, as part of the full informed consent process.

References

Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., & Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: a meta-analysis. *Psychonomic Bulletin & Review, 22*(2), 366–377. https://doi.org/10.3758/s13423-014-0699-x.

Ball, K., Berch, D. B., Helmers, K. F., Jobe, J. B., Leveck, M. D., Marsiske, M., ..., Group, A. S (2002). Effects of cognitive training interventions with older adults: a randomized controlled trial. *JAMA*, 288(18), 2271-2281.

Bandura, A., Grusec, J. E., & Menlove, F. L. (1966). Observational learning as a function of symbolization and incentive set. *Child Development, 37*(3), 499–506.

Bavelier, D., Bediou, B., & Green, C. S. (2018). Expertise and generalization: lessons from action video games. *Current Opinion in Behavioral Sciences*. https://doi.org/10.1016/j.cobeha.2018.01.012.

Bediou, B., Adams, D. M., Mayer, R. E., Tipton, E., Green, C. S., & Bavelier, D. (2018). Meta-analysis of action video game impact on perceptual, attentional. *and cognitive skills. Psychological Bulletin, 144*(1), 77–110. https://doi.org/10.1037/bul0000130.

Boot, W. R., Simons, D. J., Stothart, C., & Stutts, C. (2013). The pervasive problem with placebos in psychology: why active control groups are not sufficient to rule out placebo effects. *Perspectives on Psychological Science, 8*(4), 445–454. https://doi.org/10.1177/1745691613491271.

Büchel, C., Geuter, S., Sprenger, C., & Eippert, F. (2014). Placebo analgesia: a predictive coding perspective. *Neuron, 81*, 1223–1239. https://doi.org/10.1016/j.neuron.2014.02.042.

Colloca, L., Klinger, R., Flor, H., & Bingel, U. (2013). Placebo analgesia: psychological and neurobiological mechanisms. *Pain, 154*(4), 511–514. https://doi.org/10.1016/j.pain.2013.02.002.

Denkinger, S., Spano, L., Bingel, U., Witt, C.M., Bavelier, D. & Green, C.S. (in press). Assessing the impact of expectations in cognitive training and beyond. *Journal of Cognitive Enhancement*.

Deveau, J., Jaeggi, S. M., Zordan, V., Phung, C., & Seitz, A. R. (2014a). How to build better memory training games. *Frontiers in Systems Neuroscience, 8*, 243. https://doi.org/10.3389/fnsys.2014.00243.

Deveau, J., Ozer, D. J., & Seitz, A. R. (2014b). Improved vision and on-field performance in baseball through perceptual learning. *Current Biology, 24*(4), R146–R147. https://doi.org/10.1016/j.cub.2014.01.004.

Dweck, C. (2006). *Mindset: the new psychology of success*. New York: Random House.

Foroughi, C. K., Monfort, S. S., Paczynski, M., McKnight, P. E., & Greenwood, P. M. (2016). Placebo effects in cognitive training. *PNAS, 113*, 7470–7474. https://doi.org/10.1073/pnas.1601243113.

Green, C. S., Strobach, T., & Schubert, T. (2014). On methodological standards in training and transfer experiments. *Psychological Research, 78*(6), 756–772. https://doi.org/10.1007/s00426-013-0535-3.

Green, C. S., Bavelier, D., Kramer, A. F., Vinogradov, S., Ansorge, U., Ball, K. K., ..., Witt, C. M. (2019). Improving methodological standards in behavioral interventions for cognitive enhancement. *Journal of Cognitive Enhancement*, 3(1), 2-29.

Hrobjartsson, A., Forfang, E., Haahr, M. T., Als-Nielsen, B., & Brorson, S. (2007). Blinded trials taken to the test: an analysis of randomized clinical trials that report tests for the success of blinding. *International Journal of Epidemiology, 36*(3), 654–663. https://doi.org/10.1093/ije/dym020.

Jeffreys, H. (1961). *Theory of probability*. Oxford: Oxford University Press.

Jones, M. R., Katz, B., Buschkuehl, M., Jaeggi, S. M., & Shah, P. (2018). Exploring N-back cognitive training for children with ADHD. *Journal of Attention Disorders*. https://doi.org/10.1177/1087054718779230.

Kaptchuk, T. J., & Miller, F. G. (2015). Placebo effects in medicine. *The New England Journal of Medicine, 373*(1), 8–9. https://doi.org/10.1056/NEJMp1504023.

Karbach, J., & Unger, K. (2014). Executive control training from middle childhood to adolescence. *Frontiers in Psychology, 5*, 390. https://doi.org/10.3389/fpsyg.2014.00390.

Katz, B., Jaeggi, S. M., Buschkuehl, M., Shah, P., & Jonides, J. (2018). The effect of monetary compensation on cognitive training outcomes. *Learning and Motivation, 63*, 77–90.

Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlstrom, K., ... Westerberg, H. (2005). Computerized training of working memory in children with ADHD- a randomized, controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, 44(2), 177-186.

Kramer, A. F., Larish, J., & Strayer, D. L. (1995). Training for attentional control in dual-task settings: a comparison of young and old adults. *Journal of Experimental Psychology: Applied, 1*, 50–76.

Melby-Lervag, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology, 49*(2), 270–291.

Morey, R. D., Rouder, J. N., Jamil, T., Urbanek, S., Forner, K., & Ly, A. (2018). Computation of Bayes factors for common designs.

Nahum, M., Lee, H., & Merzenich, M. M. (2013). Principles of neuroplasticity-based rehabilitation. *Progress in Brain Research, 207*, 141–171. https://doi.org/10.1016/B978-0-444-63327-9.00009-6.

Pahor, A., Stavropoulos, T., Jaeggi, S. M., & Seitz, A. R. (2018). Validation of a matrix reasoning task for mobile devices. *Behavior Research Methods doi*. https://doi.org/10.3758/s13428-018-1152-2.

Prakash, R. S., De Leon, A. A., Patterson, B., Schirda, B. L., & Janssen, A. L. (2014). Mindfulness and the aging brain: a proposed paradigm shift. *Frontiers in Aging Neuroscience, 6*, 120. https://doi.org/10.3389/fnagi.2014.00120.

Price, D. D., Finniss, D. G., & Benedetti, F. (2008). A comprehensive review of the placebo effect: recent advances and current thought. *Annual Review of Psychology, 59*, 565–590. https://doi.org/10.1146/annurev.psych.59.113006.095941.

Rabipour, S., Morrison, C., Crompton, J., Petrucelli, M., de Oliveira Gonçalves Germano, M., Popescu, A., & Davidson, P. S. R. (2019). Few effects of a 5-week adaptive computerized cognitive training program in healthy older adults. *Journal of Cognitive Enhancement.* https://doi.org/10.1007/s41465-019-00147-2.

Rohde, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence, 35*(1), 83–92.

Schlickum, M. K., Hedman, L., Enochsson, L., Kjellin, A., & Fellander-Tsai, L. (2009). Systematic video game training in surgical novices improves performance in virtual reality endoscopic surgical simulators: a prospective randomized study. *World Journal of Surgery, 33*(11), 2360–2367.

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2010). Hundred days of cognitive training enhance broad abilities in adulthood: findings from the COGITO study. *Frontiers in Aging Neuroscience, 2*, 27.

Shipstead, Z., Hicks, K. L., & Engle, R. W. (2012a). Cogmed working memory training: does the evidence support the claims? *Journal of Applied Research in Memory and Cognition, 1*, 185–193.

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012b). Is working memory training effective? *Psychological Bulletin, 138*(4), 623–654.

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. (2016). Do "brain-training" programs work? *Psychological Science in the Public Interest, 17*(3), 103–186. https://doi.org/10.1177/1529100616661983.

Stieff, M., & Uttal, D. (2015). How much can spatial training improve STEM achievement. *Educational Psychology Review, 27*(4), 607–615.

Strobach, T., & Karbach, J. (Eds.). (2016). *Cognitive training: an overview of features and applications*. New York: Springer.

Tsai, N., Buschkuehl, M., Kamarsu, S., Shah, P., Jonides, J., & Jaeggi, S. M. (2018). (Un)Great expectations: the role of placebo effects in cognitive training. *Journal of Applied Research in Memory and Cognition, 7*(4), 564–573.

Tuttle, A. H., Tohyama, S., Ramsay, T., Kimmelman, J., Schweinhardt, P., Bennett, G. J., & Mogil, J. S. (2015). Increasing placebo responses over time in U.S. clinical trials of neuropathic pain. *Pain, 156*(12), 2616–2626. https://doi.org/10.1097/j.pain.0000000000000333.

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013a). The malleability of spatial skills: a meta-analysis of training studies. *Psychological Bulletin, 139*(2), 352–402.

Uttal, D. H., Miller, D. I., & Newcombe, N. S. (2013b). Exploring and enhancing spatial thinking: links to achievement in science, technology, engineering, and mathematics? *Current Directions in Psychological Science, 22*(5), 367–373.

Valdes, E. G., Andel, R., Lister, J. J., Gamaldo, A., & Edwards, J. D. (2017). Can cognitive speed of processing training improve everyday functioning among older adults with psychometrically defined mild cognitive impairment? *Journal of Aging and Health, 31*(4), 595–610. 898264317738828. https://doi.org/10.1177/0898264317738828.

Voudouris, N. J., Peck, C. L., & Coleman, G. (1985). Conditioned placebo responses. *Journal of Personality and Social Psychology, 48*(1), 47–53.

Wabersich, D., & Vandekerckhove, J. (2014). The RWiener package: an R package providing distribution functions for the Wiener diffusion model. *The Royal Journal, 6*(1), 49–56.

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin and Review, 14*(5), 779–804.

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E. J. (2011). Statistical evidence in experimental psychology: an empirical comparison using 855 t tests. *Perspectives on Psychological Science, 6*(3), 291–298. https://doi.org/10.1177/1745691611406923.

Wright, R., Thompson, W. L., Ganis, G., Newcombe, N. S., & Kosslyn, S. M. (2008). Training generalized spatial skills. *Psychonomic Bulletin and Review, 15*(4), 763–771.