Journal of Cognitive Enhancement
https://doi.org/10.1007/s41465-025-00316-6

REGISTERED REPORTS OR REPLICATIONS

No Expectancy Effect on Visual Attention Performance After a Single Session of Playing Sudoku: A Failed Replication of Tiraboschi et al. (2019)

Freya Joessel[1] · **Emma G. Cunningham**[1] · **C. Shawn Green**[1]

Received: 25 June 2024 / Accepted: 7 January 2025

Abstract

There is significant interest in the development of behavioral interventions to improve cognitive skills. Part of the basic-science foundation of all such endeavors is understanding the mechanisms by which cognitive performance can be altered. While most intervention studies have focused on mechanisms such as persistent load placed upon certain cognitive functions, it has been argued that participant expectations (e.g., placebo effects) may also play a role in intervention study outcomes. Researchers have thus asked whether it is possible to impact cognitive performance via expectation-mechanisms alone. In one such published study, Tiraboschi et al. (*Journal of Cognitive Enhancement 3*(4):436–444, 2019) found that participants who were given positive expectations about the impact of a (likely benign) intervention showed greater enhancements than those given no such expectations on the Useful Field of View (UFOV) visual search task. The current study sought to replicate and extend this finding using similar methods but incorporating several methodological improvements in alignment with current standards for cognitive behavioral interventions. These include improved participant and experimenter masking procedures, improved baseline group matching procedures, and the addition of a negative expectation group. Unfortunately, our results did not replicate the expectation effects found by Tiraboschi et al. (*Journal of Cognitive Enhancement 3*(4):436–444, 2019). Regardless of the information participants were provided about the impact of the intervention, all groups' UFOV performance between pre- and post-test was equivalent as assessed by Bayesian null hypothesis testing. We discuss possible reasons for this discrepancy (e.g., different populations of participants) as well as potential future directions for the field.

Keywords Useful Field of View · Expectations · Placebo · Intervention study · Replication study

Introduction

A large body of research has shown that cognitive skills such as fluid intelligence and executive functions (inhibition, updating/working memory, shifting) underpin many real-world outcomes such as academic success (Blair & Razza, 2007; Cortés Pascual et al., 2019; Ramos-Galarza et al., 2019; Titz & Karbach, 2014), professional success (Bailey, 2007; Deary et al., 2007; Rohde & Thompson, 2007), quality of life (Brown & Landgraf, 2010; Davis et al., 2010; Hsu et al., 2012), mental health (Baler & Volkow, 2006; Diamond, 2005; Jarmolowicz et al., 2013; Klingberg et al., 2002; Stordal et al., 2004; Taylor Tavares et al., 2007), physical health (Crescioni et al., 2011; Miller, 2011; Riggs et al., 2010), marital harmony (Eakin et al., 2004), and public safety (Broidy et al., 2009; Denson et al., 2011). Cognitive training offers a possible route to either bolster or remediate those cognitive skills. Indeed, ample research over the past two decades has offered promising results that suitable interventions have the potential to improve cognitive skills ranging from processing speed (Ball et al., 2007; Green & Bavelier, 2006a), to attentional control skills (Green & Bavelier, 2006b; Oei & Patterson, 2015), to spatial cognition (Spence & Feng, 2010; see Anguera et al., 2013 and Bediou et al., 2023 for a review), to working memory (Ericson & Klingberg, 2023; Klingberg, 2010; Schmiedek et al., 2019), to executive functions (Green et al., 2012; Johann & Karbach, 2021; Strobach et al., 2012), to fluid intelligence (Jaeggi et al., 2008), see (Au et al., 2015), and finally to problem solving (Shute et al., 2015).

However, outstanding methodological issues remain that may call the mechanism of action into question. Specifically,

✉ Freya Joessel
joessel@wisc.edu

[1] Department of Psychology, University of Wisconsin-Madison, 1202 W Johnson St, Madison, WI 53706, USA

Published online: 28 January 2025

while most investigators in this domain have pointed to core mechanisms inherent in training interventions themselves as the causal factor driving improved performance in cognitive task outcomes (e.g., long-term sustained load placed on various cognitive capacities; (Bavelier et al., 2010, 2012; Deveau et al., 2015; Watanabe & Sasaki, 2015), others have raised the possibility that any observed positive effects are due exclusively to expectation effects (we note that these effects are named expectation, expectancy, and placebo effects in the literature; here, we will use expectation effects; Boot et al., 2011; Melby-Lervåg et al., 2016; Sala et al., 2018; Unsworth et al., 2015; but see Green et al., 2017). Expectation effects are defined as any impact to a measured outcome that is caused, not by any intended intervention, but by the participant's belief that the activity they are engaged in will result in changes in the outcome. Such changes may be either positive (placebo) or negative (nocebo) compared to the effects of neutral or an absence of expectations.

In many domains, researchers guard against such expectation effects by ensuring that participants are unable to determine the condition to which they have been assigned. For instance, in a pharmaceutical study, it is possible to create two identical-looking pills, one containing the active ingredient and one containing no active ingredients (e.g., a sugar pill). Ideally, this study design would make it impossible for participants to determine which condition (experimental or control) they were assigned to, eliminating possible differential expectation effects across conditions (although see Denkinger et al., 2021; Green et al., 2019; Haahr & Hróbjartsson, 2006; Kolahi et al., 2009 for a discussion of the fact that this does not always work as intended).

One main reason this alternative "expectation effect-based mechanism" continues to be discussed as viable in the context of cognitive interventions is that it is more or less impossible to truly mask[1] participants to the condition to which they were assigned in such studies. Thus, there remains the potential for participants assigned to different groups (e.g., experimental versus control) to develop different expectations about the expected outcome of their experience.

Indeed, in cognitive training studies, the very features that make an intervention the experimental condition (i.e., meant to drive improvements in cognitive function) versus the control condition (i.e., not meant to drive changes in cognitive functions) will necessarily result in differences in the look and feel of the experiences. This may result in participants identifying the hypothesis of the study while participating, which in turn could influence their behavior. If this occurs, it may bias the results (Boot et al., 2011; Green et al., 2019; Rabipour & Davidson, 2015).

For example, when investigators have sought to determine whether action video game (AVG) play drives improvements in various aspects of cognition, they have typically contrasted training on an AVG with training on a game that lacks the core "action" components of AVGs (e.g., fast pacing, demanding peripheral processing under clutter, and need to make speeded decisions), but that matches the AVG in terms of experienced fun, commercial success, identification with character, amount of improvement felt, and other game qualities. In such a study, participants playing the AVG may realize that the game is attentionally demanding and thus form the expectation that this will improve their performance on certain tasks on which they were tested before the training. Such expectations may not be present when playing non-action video games that do not place heavy demands on attention, which are often used as control activities in intervention studies investigating the impact of AVGs on cognition (Bejjanki et al., 2014; Boot et al., 2008; Feng et al., 2007; Green & Bavelier, 2003, 2006b, 2006a, 2007; Novak & Tassell, 2015; Oei & Patterson, 2013; Strobach et al., 2012; Wu et al., 2012; Zhang et al., 2021).

The same general idea can be seen throughout all areas of cognitive training. For example, in a common cognitive training experiment design, a participant may be asked to complete a series of pre-test working memory tasks based on remembering a set of items, doing some kind of operation on the items, and reporting back those items in the correct order. They would then train for several hours on an intervention activity such as the N-back task, which involves remembering a shifting set of items and reporting the correct items, or a training program such as *Cogmed*, composed of gamified cognitive tasks targeting working memory. Finally, at the conclusion of training, they would complete another round of the same tests experienced at the pre-test, as a post-test. Given the similarities between the training tasks and the pre-test and post-test tasks, participants may form the expectation that the training intervention may improve their ability at post-test. Importantly, such expectations may not form when playing general knowledge games or tasks that are not cognitively challenging, which have been used as active control interventions, as playing these games does not necessarily feel like they are challenging the same aspects of cognition as the pre-test and post-test tasks. (Anguera et al., 2013; Brehmer et al., 2012; Chacko et al., 2014; Chooi & Thompson, 2012; Jaeggi et al., 2010; Klingberg et al., 2002, 2005; Lilienthal et al., 2013; Redick et al., 2013; Thorell et al., 2009).

In full intervention studies, researchers in the cognitive training domain typically attempt to combat the possibility

[1] We note that this has historically been called "blinding." However, because this practice does not involve literally making the participants or experimenter unable to see (and thus could be considered ableist), we use terms such as "masking" or "unaware," which are more accurate descriptions of the intention of experimental procedures (see Morris et al., 2007).

of such expectation-effects by describing all conditions as active (i.e., telling participants assigned to both the experimental and control conditions that their training should produce similar results) and providing some type of match between their conditions (e.g., one version of a training task that remains easy for the duration of training versus another version of the same task that becomes progressively more difficult). However, this may not always be effective due to the difficulty of creating convincing active control experiences. As such, it is crucial to understand the extent to which participant expectations effects may impact gains seen after cognitive training. This is not only important for interpreting the results of cognitive training studies as they are currently deployed, but also, if it is possible to enhance cognitive function via expectation mechanisms alone, this could be an important tool that researchers in applied settings could use to maximize the positive real-world impact of their training.

Because in true cognitive training studies researchers are typically attempting to minimize—or otherwise guard against—expectation effects, one route to examining the potential impact of expectations on cognitive training outcomes is to conduct separate studies in which expectations are deliberately manipulated and see whether these purposefully induced expectations interact with cognitive performance. As of now though, the literature supporting whether preconceived or purposefully created expectations impact cognitive training outcomes remains both rather sparse and quite mixed in outcome. For example, Vodyanyk and colleagues (2021) completed a set of experiments where participants' expectations regarding the impact of various cognitive improvement methods (working memory training, video game, "drug," mindset) were purposefully manipulated (via a combination of explicit instructions and also associative conditioning). However, none of the studies was there a significant expectation effect (Vodyanyk et al., 2021). Conversely, Parong et al. (2022) presented covertly recruited participants with material designed to induce either positive (placebo) or negative (nocebo) expectations towards cognitive training before the participants started 20 sessions of working memory training. Participants in the placebo group showed greater improvements in fluid intelligence, working memory, and task-switching abilities than those in the nocebo group. In addition, those who trained on a control general knowledge game also showed greater improvement in the presence of a positive expectation than a negative expectation.

The difference in the findings of Vodyanyk et al. (2021) and Parong et al. (2022) may be due, in part, to the duration of training. Parong et al. (2022) observed differences between placebo and nocebo expectations after many sessions of training, while Vodyanyk et al. (2021) did not observe any such differences after a single session. Yet, a recent study by Tiraboschi et al. (2019) did report that it is possible to create significant differences in performance change as a function of expectations with very short interventions. Specifically, Tiraboschi et al. (2019) had participants complete a Useful Field of View (UFOV) task (i.e., a measure of visual attention) before and after playing 15 minutes of Sudoku. Prior to the start of the experiment, the participants were randomly allocated to either a placebo or neutral condition, and an experimenter read them a text corresponding to their group. In the case of the placebo condition, the text indicated that 15 minutes of playing Sudoku would have a positive effect on spatial attention ("playing logic games such as these for a few minutes has been shown to make you think faster and be more attentive"). The control text meanwhile did not mention any possible expected impact from playing Sudoku ("you are going to play Sudoku so that you won't have to sit here and do nothing – which would be undesirable"). The authors found that the group that was given positive expectations about Sudoku improved more on the UFOV task from pre-test to post-test compared to the control group.

Because the Useful Field of View task is frequently employed in many cognitive training experiments, confirming whether expectations alone can significantly impact performance is of high importance. We thus pre-registered a replication and extension of the work by Tiraboschi et al. (2019) that sought to provide additional empirical support for this previous work while adjusting the methods to address some of the limitations of the methods present in the previous study.

The first methodological change involved the assignment of participants to groups. As is typical in the literature, Tiraboschi and colleagues (2019) randomly assigned participants to groups (e.g., placebo versus neutral). While random assignment eliminates potential biases in assignment, it also unfortunately retains the possibility that groups will be poorly matched at pre-test in terms of performance (i.e., that one group, by the random chance of assignment, will start out better/worse than the other on the outcome task). If this occurs, it is a significant threat to the interpretation of the main statistical analysis of interest (i.e., a group × time interaction; Green et al., 2013, 2019). In Tiraboschi et al. (2019), while not statistically significant, the expectation group did start numerically quite a bit worse than the control group. As such, in the case of no impact whatsoever, a group × time interaction might thus still be expected due to simple regression to the mean (or a number of other mechanisms of no interest, such as that people who start worse have more room to improve). To avoid this issue and ensure that groups were as well matched at pre-test as possible, here we assigned participants to groups using an algorithm that minimizes the variance across groups along a predefined set of variables (Sella et al., 2021). Specifically, we matched groups across gender, age, and UFOV performance at pre-test.

The second methodological change was in addressing the issue of determining the correct number of participants to recruit in order to be able to adequately answer the research question at hand. While the work of Tiraboschi et al. (2019) provides a benchmark effect size, estimating the number of participants to run (e.g., via a power analysis) typically remains a degree of investigator freedom—presenting an opportunity for bias. For example, in Tiraboschi et al. (2019), two samples of 24 and 28 participants were recruited at different time points and later combined. The group × time interaction on UFOV performance was significant in the first sample, but not in the second, and remained significant when both samples were combined, although with a higher *p*-value and a smaller effect size. Here instead, we used sequential Bayesian sampling to recruit participants until enough evidence had been collected for or against a change in group difference performance between pre-test and post-test. In other words, rather than pre-registering a number of participants based upon an effect size derived from a single study (noting that, even in the case of "true" effects, the first demonstration often comes with a larger effect size than subsequent replications), we instead pre-registered the desired strength of evidence and noted that we would continue to run participants until this level of evidence was reached. This ensured that even in the case where no difference between groups in pre-test to post-test changes were observed, we would be able to draw meaningful conclusions from the experiment (i.e., we could be sure that if we observed a null result, it would be an informative null rather than the result of a lack of conclusive evidence).

A third methodological change was implemented so as to ensure that we were examining the effects of expectation in isolation. Namely, we implemented masking procedures for both experimenters and participants (Green et al., 2019; Parong et al., 2023). By ensuring that both groups were unaware of research goals and intent, we could protect against the possibility of subtle and implicit differences in treatment influencing participant behavior. The absence of an experimenter bias effect is paramount to ensure that differences could be accurately attributed to an experimentally manipulated source.

Fourth, and finally, beyond the aforementioned adjustments to the base methodology of Tiraboschi et al. (2019), we also sought to extend the results in a few ways. One of these was to contrast the impact of positive participant expectations with negative participants' expectations. While Tiraboschi et al. (2019) contrasted positive expectations (placebo) that Sudoku would improve performance on the UFOV task with a neutral expectation control condition, we included an additional negative expectation condition (nocebo) that Sudoku playing would hinder performance on the UFOV task. Determining whether expectation effects, if observed, were the result of a push–pull mechanism (i.e., where performance could be made better or worse relative to baseline) versus a push-only mechanism (i.e., where expectations can push performance to be better but not worse) would provide important empirical evidence for future theory development with regard to the impacts of expectation effects on cognitive training outcomes. Further, we included several questionnaires to probe the self-reported beliefs about placebo/expectation effects of our participants as well as individual difference characteristics, in order to conduct exploratory analyses to remove potential confounds from our results (see the "Questionnaires" section for questionnaire descriptions).

In all, we tested the following hypotheses:

Hypothesis 1: Replication hypothesis (placebo > neutral and placebo > 0)

Participants in the placebo group will show improvements (difference between pre and post > 0) significantly greater than those in the neutral control group (note: This hypothesis does not include the additional nocebo group unique to this study).

Hypothesis 2: Extension hypothesis (placebo > neutral > nocebo)

Participants in the placebo group will improve more than the neutral group which in turn will show greater improvement than the nocebo group (note that while we phrase this in terms of "improvements," in practice, this hypothesis would also be confirmed if the neutral group shows no change and the nocebo group shows a reduction in performance from pre-test to post-test).

Materials and Methods

This study was pre-registered on OSF (https://osf.io/xfr7p).

Sample

We recruited a total 102 participants from the research participation pool at the University of Wisconsin-Madison. Participants were remunerated with extra credit for their Introduction to Psychology course. We excluded 12 participants (see reasons in the "Data Exclusion" section, yielding a final sample size of 90 participants (mean age = 19.1, SD = 0.9, range = [18;22] / 49 females, 40 males, 1 others).

We used sequential sampling with a Bayes Factor Stopping Rule, such that we would stop recruitment when the Bayes Factor for each of our hypotheses reached a predetermined threshold. Here, we used a threshold of $BF_{10} = 10$ for evidence for the alternate hypothesis and a $BF_{01} = 6$ for evidence in favor of the null, where BF_{10} represents the amount of evidence for the alternate hypothesis compared to the null, and vice versa for BF_{01}. A BF_{10} of 10 means that there is 10 times more evidence in favor of the alternate hypothesis compared to the null. The different thresholds associated with "evidence for" and "evidence against" the null are consistent with the known fact that it is more difficult to gather evidence for the null hypothesis (see discussion of Weiss, 1997 in Schönbrodt et al., 2017). Thus, in practice, we started by determining a level of evidence for the null hypothesis (we would be convinced of a null if there was $6\times$ more evidence for the null than the alternative) and then calculated an associated threshold level for the alternative hypothesis.

Our sampling involved starting by recruiting 30 participants, at which point we computed the Bayes Factor for both our hypotheses. We chose to start with 30 participants and only calculate the Bayes Factor for possible stopping in order to avoid a situation where a Bayes Factor happened to become very large by chance with a small number of participants. After the first 30 participants, if both Bayes factors fit our threshold (i.e., for Hypothesis 1 and Hypothesis 2), then we would have stopped the recruitment; otherwise, we would recompute those Bayes Factors every 3 additional participants until either both Bayes Factors reached our thresholds or our sample size reached 90 participants with usable data (30 participants per group, noting this latter value exceeded the number of participants in Tiraboschi et al., 2019).

In the end, it was the latter threshold that was reached first. We stopped recruiting participants at $N=90$ because the Bayes Factor for H1 did not cross any of our thresholds before there were 30 participants in each group. Note that, on the other hand, the Bayes Factor for H2 crossed the null hypothesis (H0) boundary ($BF10<1/6$) at $N=57$, $N=63$, and for all subsequent sample sizes (see Sect. 1 in supplementary materials).

The procedure involved in the study was approved by the University of Wisconsin–Madison Minimal Risk Research Institutional Review Board. Participants provided written informed consent to participate in the study and were debriefed on the true conditions and purpose of the study after study participation.

Design

We followed the methods described in Tiraboschi et al. (2019) as closely as possible except for where otherwise noted. The intervention component of the study was an app-based version of the classic 9×9 number-placement puzzle game Sudoku played for 15 minutes between pre-test and post-test, during which participants completed the same Useful Field of View task (Fig. 1).

We added the following changes to the design:

- As noted above, we added a third condition to the intervention (nocebo). The participants in this group were provided with text designed to induce the expectation that the intervention would have a negative impact on performance on the UFOV task (see the "Expectation Induction" section for the exact induction text used for each condition).
- As noted above, rather than using a purely random allocation scheme, we pseudo-randomized the participants into the three groups (neutral, placebo, or nocebo) on the fly during enrollment using the following procedure:

  1. We imposed that the number of males and females was the same in all groups (but the overall number of male participants may not match the number of female participants).
  2. Then, within each gender group, we used the algorithm by Sella et al. (2021) to match all the intervention groups with respect to age and overall performance at pre-test on the UFOV (i.e., hit rate across all trials).

- The presentation times of the stimulus in the UFOV were adapted for our sample such that the performance at pre-test matched the pre-test performance in the Tiraboschi et al. (2019) study (see the "Tasks" section for a justification of these changes).

Expectation Induction

The participants in the placebo group saw the following text, which is an AI-generated translation of the text presented in Tiraboschi et al. (2019) edited by the three authors:

> You will now play 15 minutes of Sudoku. Playing logic games such as these for a few minutes has been shown to make you think faster and be more attentive. We will assess your attention after playing Sudoku, and we expect improvements after you play Sudoku.

The participants in the neutral group saw the following text, which is an AI-generated translation of the text presented in Tiraboschi et al. (2019) edited by the three authors:

> You will now take a 15-min rest to allow your attention to recover.
> During this time, you are going to play Sudoku so that you won't have to sit here and do nothing– which would be undesirable.

The participants in the nocebo group saw the following text, which was generated for this study as there was no nocebo condition in the study by Tiraboschi et al. (2019):

> You will now play 15 minutes of Sudoku. Playing logic games such as these for a few minutes has been shown to drain your resources and make you less attentive.

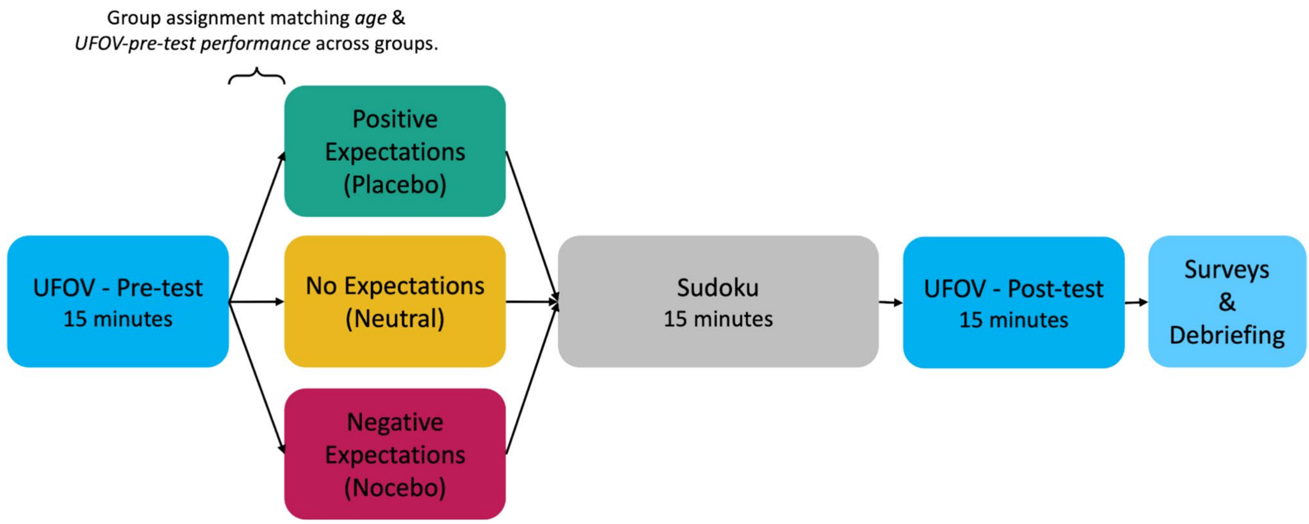

Fig. 1 Experimental design. Matching for group assignment is done according to Sella et al. (2021). Surveys include Placebo Beliefs Scale, Beliefs in expectations influencing experience, Expectation Questionnaire, and Video Game Questionnaire

> We will assess your attention after playing Sudoku, and we expect decreases in performance after you play Sudoku.

Each text was presented on a screen with a voiceover reading the text to the participant. The voices were a typical man's and woman's voice with a neutral accent for a US population. The videos presented to the participants are available on OSF (https://osf.io/5us9k/).

Tasks

The tasks included in this study were the Useful Field of View (UFOV) task which included a pre-test and a post-test before and after administration of the Sudoku "training" task.

Useful Field of View Task (Pre-Test/Post-Test) We used a task similar to the Useful Field of View (UFOV) task used in Tiraboschi et al. (2019) see Fig. 2 for the timeline of a trial). In this task, participants were briefly presented with an array of 24 objects, 3 on each of the 4 radial spokes and the 4 obliques evenly spaced, forming three "circles" located at 10°, 20°, and 30° of eccentricity. One of the items located on a radial or oblique was a target (a dark-gray filled triangle), while the remaining items were distractors (black empty squares). The participants' task was to indicate upon which of the eight spokes the target appeared. Participants completed 240 trials, 10 for each position of the target. The task lasted approximately 15 min. There was one notable difference between our version and the one from Tiraboschi et al. (2019): the presentation time of the stimulus was 20 ms when the target was located in the central circle of the stimulus, 40 ms when in the middle circle, and 50 ms when in the outer circle. We chose those values because pilot studies showed that using the presentation times reported in Tiraboschi et al. (2019) yielded performance around chance. Given the choice between exactly matching the presentation times reported by Tiraboschi et al. (2019)—that was unlikely to produce usable behavioral data based upon our pilot data—or altering the presentation times to match the behavior (accuracy) seen in Tiraboschi et al. (2019), we determined that it was better to match behavioral accuracy so that the two studies had comparable perceived difficulty for participants.

Sudoku ("Training") We use the Sudoku app by PeopleFun CG, LLC (https://apps.apple.com/us/app/sudoku/id366247306) which includes classic 9×9 Sudoku puzzles at varying levels of difficulty presented via tablet.

Questionnaires

Although Tiraboschi et al. (2019) did not include any questionnaires in their original study, we wanted to explore the extent to which individual differences in participants' a priori expectations impacted receptivity/impact to the manipulations. We thus collected several self-report measures to assess their existing beliefs about placebo effects, which were presented to the participants after they had completed the post-test UFOV task. Additionally, due to the large body of research on the impact of AVGs on UFOV performance (for a meta-analysis summarizing this literature see Bediou et al., 2023), we sought to also examine how gaming habits influence performance by administering a video game experience questionnaire. The analysis of these questionnaires

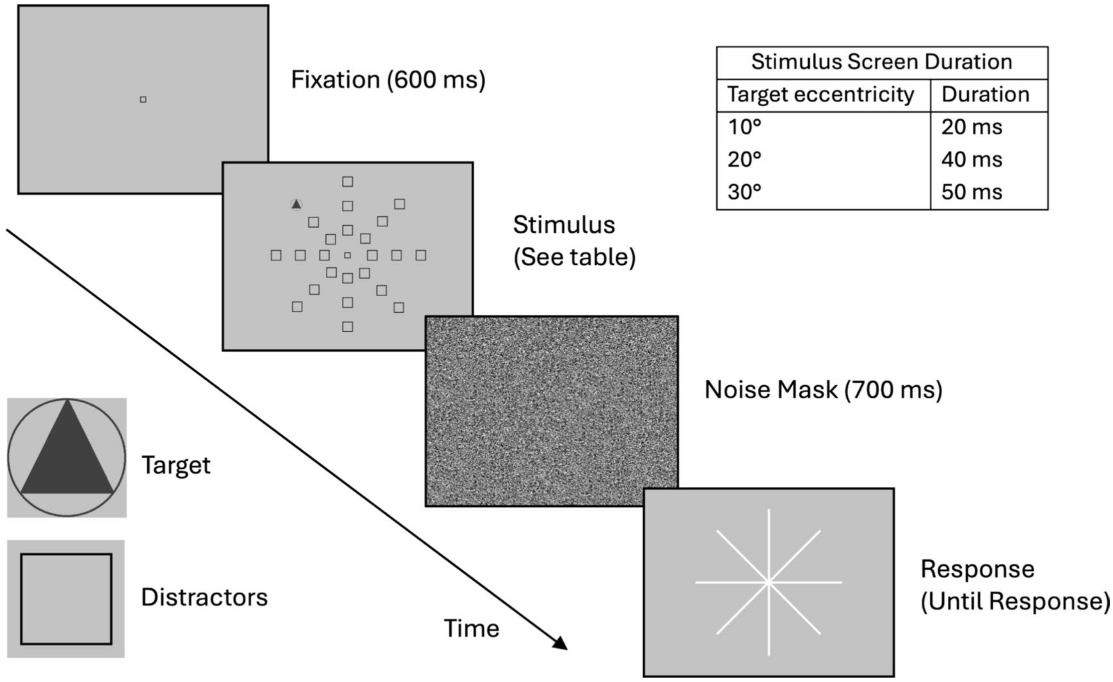

Fig. 2 Timeline of a typical trial of the UFOV task. There were a total of 240 trials per condition, with the target being presented 10 times at each of the 24 positions. To alleviate potential effects of boredom and demotivation, the participant was informed of their progression through the task at 60, 120, and 180 trials (25%, 50%, and 75% of the whole task)

are included only in our exploratory analyses because they are not part of our primary hypotheses.

Placebo Beliefs Scale (Leibowitz et al., 2019) This scale measures beliefs in placebo effects. In the medical literature, these beliefs are evaluated using two questionnaires, one from Hull et al. (2013) and another from Leibowitz et al. (2019), the latter being derived from the former. This scale is composed of the four following items scored on a Likert scale from 1 (strongly disagree) to 7 (strongly agree): "Placebo effects happen because the mind has the power to heal," "Placebo effects work because placebos influence people's expectations about a particular treatment," "Placebo effects are a part of all active medications," and "Placebo effects can occur in all illnesses and conditions."

Beliefs in Expectations Influencing Experience (Handley et al., 2009, 2013) This scale measures how much participants believed that expectations influence the outcome for experience. It is composed of the following three items scored from 0 (not at all likely) to 10 (very likely): "When learning something new.." (i) "people are likely to experience the outcome (success or failure) they expect to experience"; (ii) "if one is favorably predisposed to think they will successfully learn this new thing, one will probably succeed in learning this new thing"; and (iii) "if one is favorably predisposed to think they will fail to learn this new thing, one will probably fail in learning this new thing."

Expectation Questionnaire This questionnaire has been used previously in the intervention study literature with video games (Joessel, 2022; Zhang et al., 2021) to evaluate the participant's expectation that doing an activity (here playing Sudoku) would affect their cognition, their mood, their productivity at work, and their physical fitness. These items were rated on a scale from 1 (unsuccessful) to 5 (successful).

Video Game Questionnaire (VGQ, https://www.unige.ch/fapse/brainlearning/vgq/) This questionnaire evaluates the video gaming habits of participants over the past year and over the years before that year. The participant reported how many hours they play per week during their heaviest play period on the following categories of video games: "First/Third Person Shooter," "Action-RPG/Adventure," "Sports/Driving," "Real-Time Strategy/MOBA," "Turn-Based/Non-Action Role-Playing/Fantasy," "Turn-Based Strategy/Life Simulation/Puzzle," "Music," and "Other." The weekly hours were reported on the following scale: "never," "less than 1 hour," "between 1 and 3 hours," "between 3 and 5 hours," "between 5 and 10 hours," "more than 10 hours." Participants were sorted into four categories (Non-Video Game Players (NVGPs), Low-Tweeners, Tweeners, and Action Video Game Players (AVGPs)) as defined in the questionnaire manual.

Apparatus

Stimuli were presented on a CRT 20-inch Mitsubishi Diamond Pro 2070 SB, at a resolution of 1024 × 768 at a 100 Hz frame rate. The participant's head was set on a chinrest such that the screen was located 22 cm from their eyes.

Masking (See Footnote 1)

In a departure from the methods of Tiraboschi et al. (2019), our study included masking procedures to ensure that both participants and experimenters were masked to the assigned group of the participant. Expectation induction was performed using a video that the experimenters were not aware of, and the videos were designed in such a way that the same basic video was shown to all participants. These were automatically presented to the participant at the end of the pre-test task in the form of a PowerPoint slide with a voice over text. The voice was either a man's or a woman's, and the voice's gender was fully counterbalanced across participant gender and induction group in order to eliminate any possible effects of gender on participant trust in the instructions (see the discussion of experimenter × participant gender interaction effect on study outcome in Nichols and Maner (2008). Because the experiment involved deception, consistent with best-ethical practices, at the very end of the experiment, participant debriefing was performed by one of two experimenters that were made aware of the condition of the participant through a text message sent to their phone. The masked experimenters running the participants were not aware of this and the two unmasked experimenters only had contact with participants during the debriefing. This debriefing was performed for all participants, even the ones in the neutral group that did not receive any deceitful information, such that the experimenters in contact with the participants would remain unaware that not all participants were in the same condition.

Procedure

Our procedure closely matched the procedure from Tiraboschi et al. (2019), with a few key differences highlighted below. Participants first gave informed consent. Then they completed the first session of the UFOV task (pre-test) where they were given written instructions and eight practice trials before starting with the official task (unlike in Tiraboschi et al. (2019) where participants only saw the written instructions but no practice trials). If they reported being unable to even perceive the stimuli during these eight trials, we offered them the opportunity to do the practice once more before moving on to the assessment part of the task. At the end of the task, their performance on the UFOV was evaluated as the percentage of correct responses over all the trials. This performance was used in conjunction with their gender (which we asked participants to provide at the beginning of the task; note: this information was hidden from the experimenter) to assign the participants in groups that were matched in gender and UFOV performance at pre-test using the algorithm from Sella et al. (2021). To mask the experimenter to the condition of the participant, the induction text was automatically presented to the participant at the end of the task via text and accompanying dictation. This is unlike Tiraboschi et al. (2019) where participants were randomly assigned to their group and had the experimenter read the induction text aloud to the participant before pre-test.

The participants then played Sudoku for 15 minutes on an iPad on an easy level. Participants who did not know how to play Sudoku were instructed on how to play the game, and all participants were asked to start a new game if they finished their first one before the 15 minutes were over. Participants completed on average 2 Sudoku games (range [1; 5]).

Once the 15 minutes were over, the participants completed the second session of the UFOV (post-test), which was identical to the pre-test. After that, the participants completed the questionnaires (demographic information, beliefs in placebo effects, beliefs in expectancy effects, expectation that Sudoku had any impact on their cognition, and video game questionnaire). Once they were done, all participants went through debriefing with a second experimenter disclosing any deception in the experiment and answering any questions they had. The second experimenter was informed of the participant's condition through a text sent directly to their smartphone. The first experimenter left the room for this debriefing, thus staying masked to the fact that this experiment involved an expectation manipulation component. As far as we know, the questionnaires, experimenter masking, and debriefing were not part of the Tiraboschi et al. (2019) study.

All the instructions given to the participants are reported in the OSF repository for this study (osf.io/5us9k.).

Analysis Plan

Statistical Models

To test Hypothesis 1 (placebo group > neutral group), we evaluated the effect of expectation between the placebo and neutral groups by performing a one-way ANCOVA on post-test performance on the UFOV task with pre-test performance on the UFOV as a covariate and eccentricity (10°, 20°, or 30°) as a within-subject fixed factor.

We will consider that Hypothesis 1 is supported if the ANCOVA reveals enough evidence for a difference between groups and that post-test performance (while controlling for pre-test performance) is higher for the placebo group compared to the neutral group.

To test Hypothesis 2, we will evaluate the effect of expectation between the placebo, neutral, and nocebo groups by performing a one-way ANCOVA on post-test performance on the UFOV task with pre-test performance on the UFOV as a covariate and eccentricity (10°, 20°, or 30°) as a within-subject fixed factor.

We will consider that Hypothesis 2 is supported if the ANCOVA reveals enough evidence for a difference between groups and that post-test performance (while controlling for pre-test performance) is (i) higher for the placebo group than the neutral group and (ii) higher for the neutral group than the nocebo group.

All Bayesian analyses were conducted using JASP (JASP Team, 2024). In addition, a small guide on how to interpret results from Bayesian ANOVAs and *t*-tests can be found in Sect. 3 of the supplementary materials.

Data Exclusion

Participants that did not comply with the task were excluded from the analyses. This was monitored using the distribution of the number of times the participant used each response key. If two keys constitute more than 60% of the total strokes, then that participant was excluded.

We also excluded participants whose performance was below a threshold of chance level + 5% (which is 12.5% + 5% = 17.5%) for more than 4 of the 6 within subject conditions (the 3 eccentricities at pre-test and post-test).

We also excluded any participant with missing data for the UFOV task. Participants with missing data on the questionnaires were still included in the main analyses pertaining to UFOV performance.

Exploratory Analyses

As part of our pre-registration, we also briefly discussed several exploratory analyses that we would conduct, noting that these were either not strongly hypothesis driven (i.e., our hope was for them to point the way to future, more-confirmatory research) or else that it was unclear whether our sampling would allow for a strong test of any ideas (e.g., because we did not recruit based upon video game play, it was unclear whether or not our sample would include sufficient spread in gaming activity to provide useful information). We note these analyses below for completeness with our pre-registration. The results can be found in Sect. 4 of the supplementary materials.

Exploratory Analysis 1: We performed a Bayesian repeated measure ANOVA with group (neutral, placebo, nocebo) as a between subject factor, and eccentricity (10°, 20°, or 30°) and time (pre-test, post-test) as within-subject factors. The goal of this analysis was to investigate the difference in statistical power with ANCOVAs and RM-ANOVAs in intervention study statistical analyses.

Exploratory Analysis 2: We also investigated the potential influence of self-reported a priori expectations and beliefs in placebo effects on the potential score difference between pre-test and post-test.

Exploratory Analysis 3: As part of ongoing research regarding the influence of video games on visual attention, we also examined how gaming habits influence performance on the UFOV at pre-test, as well the learning rate between the UFOV pre- and post-tests.

Finally, we conducted one additional exploratory analysis (Exploratory Analysis 4) that was not pre-registered examining potential interactions between participant gender and the gender of the voice reading the expectation text.

Results

In the following sections, we present the results of our confirmatory analyses (Hypotheses 1 and 2).

Main Analyses

Data for the main analyses are shown in Fig. 3.

Hypothesis 1

The results of the one-way ANCOVA on post-test performance on the UFOV with pre-test performance on the UFOV as a covariate, eccentricity (10°, 20°, or 30°) as a within-subject fixed factor and group (placebo or neutral) as between-subject fixed factor showed that there was more than 10^{14} times more evidence for a main effect of session (i.e., pre-test vs post-test) than against this effect, 3.14 times more evidence for a main effect of eccentricity than against this effect, 6.06 times (= 1/0.165) more evidence against a main effect of group than for this effect, and 12.3 (= 0.081) times more evidence against the eccentricity × group interaction than for this interaction. There was no conclusive evidence for or against any of the other interactions. The full table of models and associated Bayes Factors can be found in Sect. 2.1 of the supplementary materials.

The primary result of interest is the fact that we found 6.06 × more evidence against Hypothesis 1 than for it. In other words, the evidence strongly pointed toward a null result, wherein there was no difference in the placebo group as compared to the neutral control with respect to the critical measure (i.e., post-test performance after controlling for the various relevant factors). This indicates that expectation effects did not cause improvements in UFOV performance in the placebo group when compared to the neutral group.

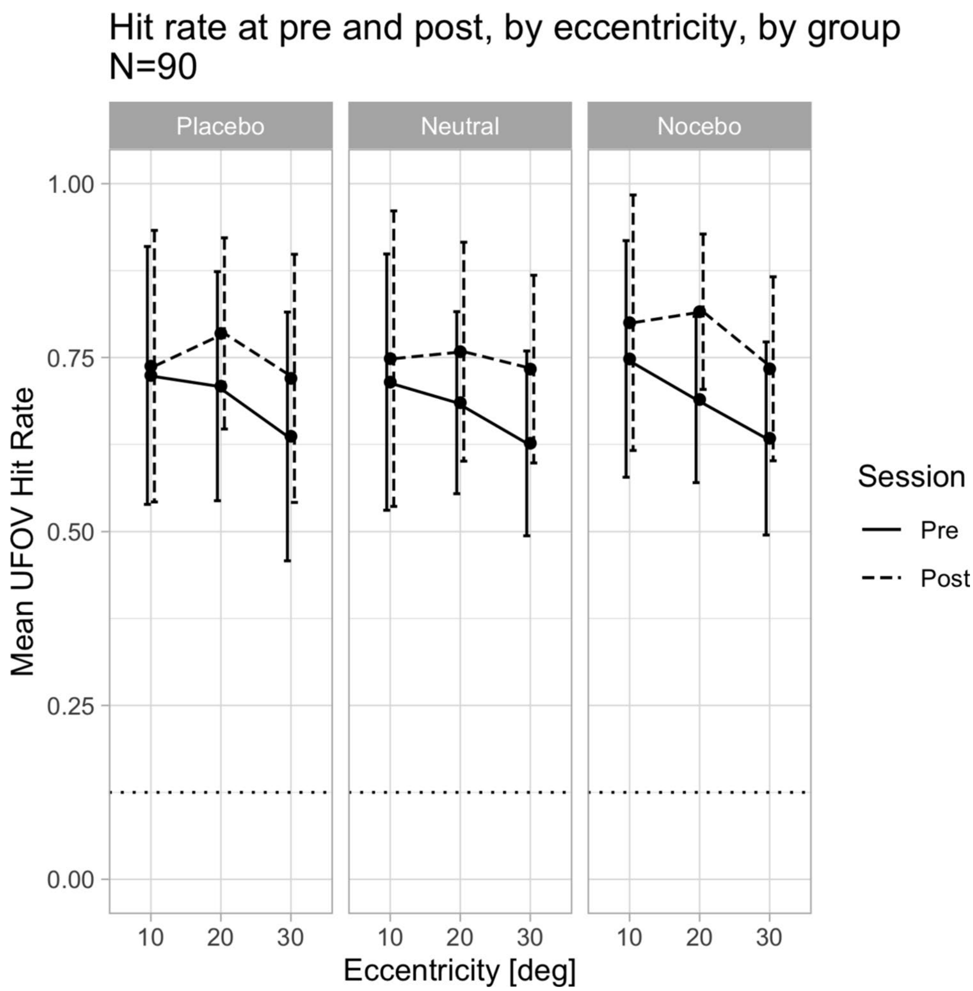

Fig. 3 Mean hit rate of each group on the UFOV task for each eccentricity and each session. Black horizontal dotted lines represent chance level (12.5%). Only the main effect of session was significant. Error bars represent standard deviations

Hypothesis 2

The results of the one-way ANCOVA on post-test performance on the UFOV with pre-test performance on the UFOV as a covariate, eccentricity (10°, 20°, or 30°) as a within-subject fixed factor and group (placebo, neutral, or nocebo) as between-subject fixed factor showed that there was more than 10^{14} times more evidence for a main effect of session (i.e., pre-test vs post-test) than against this effect, 5.83 times more evidence for a main effect of eccentricity than against this effect, 9.62 (= 1/0.104) times more evidence against a main effect of group than for this effect, and 83.3 (= 0.012) times more evidence against the eccentricity × group interaction than for this interaction. There was no conclusive evidence for or against any of the other interactions. The full table of models and associated Bayes Factors can be found in Sect. 2.3 the supplementary materials.

Similar to the analysis examining Hypothesis 1, the main result of interest is the fact that we found, in this case, 9.62 times more evidence against Hypothesis 2 than for it. In other words, the evidence strongly pointed toward a null result, wherein there was no difference across the three groups (placebo, neutral control, nocebo) with respect to the critical measure (i.e., post-test performance after controlling for the various relevant factors). This indicates that expectation effects did not cause improvements in UFOV performance in the placebo group when compared to the neutral group and when compared to the nocebo group.

Discussion

We set out to replicate and extend the UFOV portion of the study from Tiraboschi et al. (2019) wherein participants who were given positive expectations about the effect of playing Sudoku on their performance on the UFOV task improved significantly more from pre-test to post-test as compared to a control group who were given neutral expectations about the effect of Sudoku. Unfortunately, the replication was not successful, with there being 6 × more evidence for a null than the alternative hypothesis (Hypothesis 1). Perhaps not surprisingly then, our extension to Tiraboschi et al. (2019) also produced a null result (in this case with over 9 × more evidence for the null than the alternative Hypothesis 2).

While inconsistent with the results of Tiraboschi et al. (2019), our results do fit within much of the recent literature utilizing very short durations of "training" in expectation-induction type studies which also failed to find significant expectation effects. For instance, Vodyanyk and colleagues (2021) failed to find an expectation effect when utilizing several different types of very short term "training" (working memory training, video game, "drug," mindset) accompanied by expectation-induction protocols. Similarly, Tsai et al. (2018) observed no difference in the impact of seven short sessions of training (either active N-back training or control trivia training) as a function of different expectation-inducing narrations participants received at the start of training (one placebo and one neutral).

Critically, we note that this does not necessarily mean that no such effects might be observed with longer training durations. Indeed, changes in cognitive and perceptual abilities often require at least several hours of training spread across several weeks to produce an effect on cognition, with increasing reliability as training dose increases (Bediou et al., 2023, p. 202; Schwaighofer et al., 2015; Stepankova et al., 2014). In addition, participants experience learning in their everyday life and may have strong lived experiences that 15 minutes on any task is not enough to produce increased cognitive performance. This is important because previous studies on placebo effects in pain perception have shown that lived experiences tend to create much stronger placebo effects than explicit expectations (Benedetti et al., 2003; Colloca & Benedetti, 2006; Zunhammer et al., 2017). In the context of cognitive training, this might also mean that expectations effects may not be elicited using explicit expectation induction alone, but only when such inductions are presented in conjunction with an associative learning (or conditioning) experience where the participant's expectations are confirmed through a lived experience (e.g., the participant completing an easier version of the pre-test task(s) at post-test while believing that it is the same task as at pre-test and thus experiencing perceived task improvement). It is very likely however that such conditioning (or associative learning) may only be effective in the context of longer training interventions. This might explain why significant expectation results were not observed in the longer duration (seven 25-min sessions) training study from Tsai et al. (2018) that only used explicit expectations nor in the single-session interventions from Vodyanyk and colleagues (2021) where both explicit expectations and associative learning were used, but such effects were observed in the training study employed by Parong et al. (2022), which had a longer duration (twenty 20-min sessions) and an associative learning experience mid-way through the training. In addition, we note that there is evidence that people who enter a study with strong beliefs might be particularly susceptible to placebo effects (see Foroughi et al., 2016). Further studies are needed to clarify the role of a-priori participant beliefs, different expectation induction schemes, and intervention duration.

Still, such a view would fit nicely in the framework suggesting that expectations may be understood within a Bayesian cue combination (Denkinger et al., 2021). In such a framework, expectations act as priors on the outcome of a given experience. For example, in the context of pain perception, strong expectations that a cream has an analgesic effect will be combined with the actual perceived pain to produce pain level reports that are lower than what the participant would feel without the cream, independent of whether the cream has an actual analgesic effect (Büchel et al., 2014). In the context of cognitive training, this can be understood as expectations having a multiplicative effect rather than an additive effect on cognitive improvements. In other words, expectations regarding the effect of a cognitive training will modulate the effect that this training has on cognition, but if that effect is limited or non-existent to begin with, then expectations will not yield any cognitive improvements (which was the case in Parong et al., 2022, where the impact of expectations was added "on top" of the impact of the active working memory training).

While no study is without limitations, we believe that our methods were, if anything, more robust than those previously employed in Tiraboschi et al. (2019). Changes made to the original design include pseudo-random group-assignment to ensure a strong match in UFOV pre-test performance (thus ensuring that post-test results were clearly interpretable), strong masking of experimental intent for both participants and experimenters (instead of no masking procedure), and a slightly larger sample with $N=30$ per group, giving 90 participants in total (instead of 26 per group, 52 participants in total) that arose via a pre-registered stopping rule.

We note that our presentation duration differed from that reported by Tiraboschi et al. (2019). They reported performance around 75% with display durations of 10 ms for targets at 10° of eccentricity and 20 ms for targets and 20° and 30° of eccentricity utilizing E-Prime Software. However, our own pilots as well as previous experiments suggest that at such durations, performance tends to be around 35% (Green & Bavelier, 2003). As such, we were forced to make a choice between matching Tiraboschi et al.'s (2019) reported participant behavior or their reported presentation time. We opted to match the reported participant behavior based on the belief that in the context of pre-/post-test expectation effects, the experience of the participant (i.e., their experienced difficulty) would likely matter more than the parameters in the software.

This discrepancy in performance might be due to the software used to present the stimuli. We used Psychtoolbox, a widely used MATLAB toolbox for psychophysics experiments, to measure the exact duration of presentation of the stimulus. Out of 44,160 stimulus presentations in our experiment, only 6 did not match the specified presentation time, and thus, we feel quite confident regarding our timings. However, while we attempted to verify and match contrast levels (e.g., making use of the code generously provided by Tiraboschi et al., 2019), differences in contrast could have impacted final performance. We feel it is unlikely though that this would have then had the knock-on effect of specifically disrupting an expectation effect, given that our overall performance behavior did end up largely (though not perfectly) matching that of Tiraboschi et al. (2019).

Another difference between the two studies is the mean age of the samples; our sample had a mean age of 19.1 years (SD = 0.9), while the sample in Tiraboschi et al. (2019) had a mean age of 23.6 years (SD = 3.6). Many models highlight the importance of lived experiences in the induction of placebo effects (which vary greatly with age). However, the empirical data on the extent to which age is an important predictor of placebo effects is still somewhat unclear (Geers et al., 2019). In our study, expectations were presented as arising from scientific evidence (an authority figure). As such, we might have a priori expected, if anything, that a younger sample (as in our study) would show a stronger effect than an older sample. However, it would be useful for future work to further explore this issue.

A last key difference between our study and the previous work by Tiraboschi et al. (2019) was in the location of the study. Our study was conducted in a North American university with a predominantly white sample, while the study from Tiraboschi et al. (2019) was conducted in a Brazilian university, where the sample population is culturally and ethnically different from that in the Midwestern United States. This could then reflect an interesting cultural difference to explore in future work.

Finally, it is worth discussing in more depth the success of two of the methodological advances utilized here, as they could be valuable to the field going forward—namely (i) using a matching algorithm on a continuous variable with a priori unknown distribution for our pseudo-randomized assignment at pre-test (Sella et al., 2021) and (ii) a Bayesian Sequential Sampling Procedure (Schönbrodt et al., 2017), which alleviated the need to do a power analysis on an unclear effect size, which reduces the chances of making errors in our inferences without justifying this number based on experiments that were not necessarily the same. Regarding the group matching procedure, results showed that overall groups (males and females combined) were well-matched, and permutation analyses showed that the groups were better matched on UFOV pre-test performance than 83.4% of 1000 permutations, i.e., better than 83.4% of possible samples obtained through purely random group assignment. This is encouraging for future intervention studies with designs including more than two experimental arms and where matching may be required on more than one variable (e.g., age, growth mindset endorsement; for an example, see Joessel, 2022). This is particularly true given that the matching procedure works on an "on-the-fly" basis (i.e., as participants are enrolling and being pre-tested), as this is the norm in cognitive training interventions. While the use of such procedures has been argued for in the cognitive training domain for over a decade (e.g., Green et al., 2013), there are, to-date, few-to-no clear demonstrations of their utility of the type provided here.

The Bayesian sequential sampling method meanwhile showed that the Bayes Factor remains close to 1 (low evidence in either direction) for samples with fewer than 60 participants and that it may vary quite considerably from the addition of only 3 participants, even as far as showing evidence for H1 at a low sample size before decreasing to show consistent evidence for H0 at higher sample sizes. We note that for both of these procedures to work best, it is crucial to immediately remove any to-be-excluded participants (i.e., participants who were non-compliant, or whose performance was at chance) as early as possible in the sequential sampling as their presence could create issues for both the matching algorithm and the sequential sampling procedure (e.g., with respect to the matching procedure, it may lead the algorithm to correct for variance that will be absent from the final sample, thereby creating undesirable group imbalances, and with respect to the stopping rule, they may cause an erroneous stop to data collection). We suggest that future work should utilize and wherever possible continue to build upon these two methodological aspects as they hold the potential to greatly aid the extent to which data produced by the field is robust, unbiased, and fully interpretable.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s41465-025-00316-6.

Funding Partial financial support for the following research was received from NIA Grant R01AG076157 and Office of Naval Research Grant N00014-22–1-2283 to CSG.

Data Availability This study was pre-registered on OSF (https://osf.io/xfr7p). Data and materials for this study can be found here https://osf.io/5us9k/.

Declarations

Ethical Approval The procedure involved in the study was approved by the University of Wisconsin–Madison Minimal Risk Research Institutional Review Board. Participants provided written informed consent to participate in the study and were debriefed on the true conditions and purpose of the study after study participation.

Conflict of Interest One of the authors of this paper is Editor-in-Chief of the journal. This paper was handled by a Senior Editorial Board member who assumed responsibility for its peer review. The Editor-in-Chief was not involved in the peer review process. The other authors have no competing interests to declare that are relevant to the content of this article.

Trial Registration Information - Open Science Framework registration DOI: https://doi.org/10.17605/OSF.IO/XFR7P.

- Date of registration: January 9, 2024.
- Internet archive link: https://archive.org/details/osf-registrations-xfr7p-v1.

References

Anguera, J. A., Boccanfuso, J., Rintoul, J. L., Al-Hashimi, O., Faraji, F., Janowich, J., Kong, E., Larraburo, Y., Rolle, C., Johnston, E., & Gazzaley, A. (2013). Video game training enhances cognitive control in older adults. *Nature, 501*(7465), 97–101. https://doi.org/10.1038/nature12486

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

Bailey, C. E. (2007). Cognitive accuracy and intelligent executive function in the brain and in business.*Annals of the New York Academy of Sciences, 1118*(1), 122–141. https://doi.org/10.1196/annals.1412.011

Baler, R. D., & Volkow, N. D. (2006). Drug addiction: The neurobiology of disrupted self-control. *Trends in Molecular Medicine, 12*(12), 559–566. https://doi.org/10.1016/j.molmed.2006.10.005

Ball, K., Edwards, J. D., & Ross, L. A. (2007). The impact of speed of processing training on cognitive and everyday functions. *The Journals of Gerontology: Series B*, *62*(Special_Issue_1), 19–31. https://doi.org/10.1093/geronb/62.special_issue_1.19

Bavelier, D., Levi, D. M., Li, R. W., Dan, Y., & Hensch, T. K. (2010). Removing brakes on adult brain plasticity: From molecular to behavioral interventions. *Journal of Neuroscience, 30*(45), 14964–14971. https://doi.org/10.1523/JNEUROSCI.4812-10.2010

Bavelier, D., Green, C. S., Pouget, A., & Schrater, P. (2012). Brain plasticity through the life span: Learning to learn and action video games. *Annual Review of Neuroscience, 35*(1), 391–416. https://doi.org/10.1146/annurev-neuro-060909-152832

Bediou, B., Rodgers, M. A., Tipton, E., Mayer, R. E., Green, C. S., & Bavelier, D. (2023). Effects of action video game play on cognitive skills: A meta-analysis. *Technology, Mind, and Behavior*, *4*(1). https://doi.org/10.1037/tmb0000102

Bejjanki, V. R., Zhang, R., Li, R., Pouget, A., Green, C. S., Lu, Z.-L., & Bavelier, D. (2014). Action video game play facilitates the development of better perceptual templates. *Proceedings of the National Academy of Sciences, 111*(47), 16961–16966. https://doi.org/10.1073/pnas.1417056111

Benedetti, F., Pollo, A., Lopiano, L., Lanotte, M., Vighetti, S., & Rainero, I. (2003). Conscious expectation and unconscious conditioning in analgesic, motor, and hormonal placebo/nocebo responses. *The Journal of Neuroscience, 23*(10), 4315–4323. https://doi.org/10.1523/JNEUROSCI.23-10-04315.2003

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development, 78*(2), 647–663. https://doi.org/10.1111/j.1467-8624.2007.01019.x

Boot, W. R., Kramer, A. F., Simons, D. J., Fabiani, M., & Gratton, G. (2008). The effects of video game playing on attention, memory, and executive control. *Acta Psychologica, 129*(3), 387–398. https://doi.org/10.1016/j.actpsy.2008.09.005

Boot, W. R., Blakely, D. P., & Simons, D. J. (2011). Do action video games improve perception and cognition? *Frontiers in Psychology, 2*, 226. https://doi.org/10.3389/fpsyg.2011.00226

Brehmer, Y., Westerberg, H., & Bäckman, L. (2012). Working-memory training in younger and older adults: Training gains, transfer, and maintenance. *Frontiers in Human Neuroscience, 6*, 63. https://doi.org/10.3389/fnhum.2012.00063

Broidy, L. M., Tremblay, R. E., Brame, B., Fergusson, D., Laird, R., Moffitt, T. E., Nagin, D. S., Bates, J. E., Loeber, R., Lynam, D. R., Pettit, G. S., & Vitaro, F. (2009). Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: A six-site, cross-national study. *Developmental psychology, 39*(2), 222.

Brown, T. E., & Landgraf, J. M. (2010). Improvements in executive function correlate with enhanced performance and functioning and health-related quality of life: Evidence from 2 large, double-blind, randomized, placebo-controlled trials in ADHD. P*ostgraduate Medicine, 122*(5), 42–51.https://doi.org/10.3810/pgm.2010.09.2200

Büchel, C., Geuter, S., Sprenger, C., & Eippert, F. (2014). Placebo analgesia: A predictive coding perspective. *Neuron, 81*(6), 1223–1239. https://doi.org/10.1016/j.neuron.2014.02.042

Chacko, A., Bedard, A. C., Marks, D. J., Feirsen, N., Uderman, J. Z., Chimiklis, A., Rajwan, E., Cornwell, M., Anderson, L., Zwilling, A., & Ramon, M. (2014). A randomized clinical trial of Cogmed working memory training in school-age children with ADHD: A replication in a diverse sample using a control condition. *Journal of Child Psychology and Psychiatry, 55*(3), 247–255. https://doi.org/10.1111/jcpp.12146

Chooi, W.-T., & Thompson, L. A. (2012). Working memory training does not improve intelligence in healthy young adults. *Intelligence, 40*(6), 531–542. https://doi.org/10.1016/j.intell.2012.07.004

Colloca, L., & Benedetti, F. (2006). How prior experience shapes placebo analgesia. *Pain, 124*(1), 126–133. https://doi.org/10.1016/j.pain.2006.04.005

Cortés Pascual, A., Moyano Muñoz, N., & Quílez Robres, A. (2019). The relationship between executive functions and academic performance in primary education: Review and meta-analysis. *Frontiers in Psychology, 10*, 1582. https://doi.org/10.3389/fpsyg.2019.01582

Crescioni, A. W., Ehrlinger, J., Alquist, J. L., Conlon, K. E., Baumeister, R. F., Schatschneider, C., & Dutton, G. R. (2011). High trait self-control predicts positive health behaviors and success in weight loss. *Journal of Health Psychology, 16*(5), 750–759. https://doi.org/10.1177/1359105310390247

Davis, J. C., Marra, C. A., Najafzadeh, M., & Liu-Ambrose, T. (2010). The independent contribution of executive functions to health related quality of life in older women. *BMC Geriatrics, 10*(1), 16.https://doi.org/10.1186/1471-2318-10-16

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement.*Intelligence, 35*(1), 13–21. https://doi.org/10.1016/j.intell.2006.02.001

Denkinger, S., Spano, L., Bingel, U., Witt, C. M., Bavelier, D., & Green, C. S. (2021). Assessing the impact of expectations in cognitive training and beyond. *Journal of Cognitive Enhancement, 5*(4), 502–518. https://doi.org/10.1007/s41465-021-00206-7

Denson, T. F., Pedersen, W. C., Friese, M., Hahm, A., & Roberts, L. (2011). Understanding impulsive aggression: Angry rumination and reduced self-control capacity are mechanisms underlying the provocation-aggression relationship. *Personality and Social Psychology Bulletin, 37*(6), 850–862.https://doi.org/10.1177/0146167211401420

Deveau, J., Jaeggi, S. M., Zordan, V., Phung, C., & Seitz, A. R. (2015). How to build better memory training games. *Frontiers in Systems Neuroscience, 8*, 243. https://doi.org/10.3389/fnsys.2014.00243

Diamond, A. (2005). Attention-deficit disorder (attention-deficit/hyperactivity disorder without hyperactivity): A neurobiologically and behaviorally distinct disorder from attention-deficit/hyperactivity disorder (with hyperactivity). *Development and Psychopathology, 17*(03).https://doi.org/10.1017/S0954579405050388

Eakin, L., Minde, K., Hechtman, L., Ochs, E., Krane, E., Bouffard, R., Greenfield, B., & Looper, K. (2004). The marital and family functioning of adults with ADHD and their spouses. *Journal of Attention Disorders, 8*(1), 1–10. https://doi.org/10.1177/108705470400800101

Ericson, J., & Klingberg, T. (2023). A dual-process model for cognitive training. *Npj Science of Learning, 8*(1), 12. https://doi.org/10.1038/s41539-023-00161-2

Feng, J., Spence, I., & Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychological Science, 18*(10), 850–855. https://doi.org/10.1111/j.1467-9280.2007.01990.x

Foroughi, C. K., Monfort, S. S., Paczynski, M., McKnight, P. E., & Greenwood, P. M. (2016). Placebo effects in cognitive training. *Proceedings of the National Academy of Sciences, 113*(27), 7470–7474. https://doi.org/10.1073/pnas.1601243113

Geers, A. L., Briñol, P., & Petty, R. E. (2019). An analysis of the basic processes of formation and change of placebo expectations. *Review of General Psychology, 23*(2), 211–229. https://doi.org/10.1037/gpr0000171

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature, 423*(6939), 534–537. https://doi.org/10.1038/nature01647

Green, C. S., & Bavelier, D. (2006a). Effect of action video games on the spatial distribution of visuospatial attention. *Journal of Experimental Psychology: Human Perception and Performance, 32*(6), 1465–1478. https://doi.org/10.1037/0096-1523.32.6.1465

Green, C. S., & Bavelier, D. (2006b). Enumeration versus multiple object tracking: The case of action video game players. *Cognition, 101*(1), 217–245. https://doi.org/10.1016/j.cognition.2005.10.004

Green, C. S., & Bavelier, D. (2007). Action-video-game experience alters the spatial resolution of vision. *Psychological Science, 18*(1), 88–94. https://doi.org/10.1111/j.1467-9280.2007.01853.x

Green, C. S., Sugarman, M. A., Medford, K., Klobusicky, E., & Bavelier, D. (2012). The effect of action video game experience on task-switching. *Computers in Human Behavior, 28*(3), 984–994. https://doi.org/10.1016/j.chb.2011.12.020

Green, C. S., Strobach, T., & Schubert, T. (2013). On methodological standards in training and transfer experiments. *Psychological Research Psychologische Forschung, 78*(6), 756–772. https://doi.org/10.1007/s00426-013-0535-3

Green, C. S., Kattner, F., Eichenbaum, A., Bediou, B., Adams, D. M., Mayer, R. E., & Bavelier, D. (2017). Playing some video games but not others is related to cognitive abilities: A critique of Unsworth et al. (2015). *Psychological Science, 28*(5), 679–682. https://doi.org/10.1177/0956797616644837

Green, C. S., Bavelier, D., Kramer, A. F., Vinogradov, S., Ansorge, U., Ball, K. K., Bingel, U., Chein, J. M., Colzato, L. S., Edwards, J. D., Facoetti, A., Gazzaley, A., Gathercole, S. E., Ghisletta, P., Gori, S., Granic, I., Hillman, C. H., Hommel, B., Jaeggi, S. M., … Witt, C. M. (2019). Improving methodological standards in behavioral interventions for cognitive enhancement. *Journal of Cognitive Enhancement*, *3*(1), 2–29. https://doi.org/10.1007/s41465-018-0115-y

Haahr, M. T., & Hróbjartsson, A. (2006). Who is blinded in randomized clinical trials? A study of 200 trials and a survey of authors. *Clinical Trials, 3*(4), 360–365. https://doi.org/10.1177/1740774506069153

Handley, I. M., Albarracín, D., Brown, R. D., Li, H., Kumkale, E. C., & Kumkale, G. T. (2009). When the expectations from a message will not be realized: Naïve theories can eliminate expectation–congruent judgments via correction. *Journal of Experimental Social Psychology, 45*(4), 933–939. https://doi.org/10.1016/j.jesp.2009.05.010

Handley, I. M., Fowler, S. L., Rasinski, H. M., Helfer, S. G., & Geers, A. L. (2013). Beliefs about expectations moderate the influence of expectations on pain perception. *International Journal of Behavioral Medicine, 20*(1), 52–58. https://doi.org/10.1007/s12529-011-9203-4

Hsu, C. L., Nagamatsu, L. S., Davis, J. C., & Liu-Ambrose, T. (2012). Examining the relationship between specific cognitive processes and falls risk in older adults: A systematic review. *Osteoporosis International, 23*(10), 2409–2424. https://doi.org/10.1007/s00198-012-1992-z

Hull, S. C., Colloca, L., Avins, A., Gordon, N. P., Somkin, C. P., Kaptchuk, T. J., & Miller, F. G. (2013). Patients' attitudes about the use of placebo treatments: Telephone survey. *BMJ, 347*, f3757–f3757. https://doi.org/10.1136/bmj.f3757

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences, 105*(19), 6829–6833. https://doi.org/10.1073/pnas.0801268105

Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y.-F., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning—Implications for training and transfer. *Intelligence, 12*, 625–635.

Jarmolowicz, D. P., Mueller, E. T., Koffarnus, M. N., Carter, A. E., Gatchalian, K. M., & Bickel, W. K. (2013). Executive dysfunction in addiction. In J. MacKillop & H. de Wit (Eds.), *The Wiley-Blackwell Handbook of Addiction Psychopharmacology* (pp. 27–61). Wiley-Blackwell. https://doi.org/10.1002/9781118384404.ch2

JASP Team. (2024). *JASP* (Version 0.18.3) [Computer software].

Joessel, F. (2022). *Development of a video game to investigate the AVG features promoting attentional control* [Université de Genève]. https://doi.org/10.13097/ARCHIVE-OUVERTE/UNIGE:164866

Johann, V. E., & Karbach, J. (2021). Educational application of cognitive training. In T. Strobach & J. Karbach (Eds.), *Cognitive Training* (pp. 333–350). Springer International Publishing. https://doi.org/10.1007/978-3-030-39292-5_23

Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences, 14*(7), 317–324. https://doi.org/10.1016/j.tics.2010.05.002

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology, 24*(6), 781–791. https://doi.org/10.1076/jcen.24.6.781.8395

Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlström, K., Gillberg, C. G., Forssberg, H., & Westerberg, H. (2005). Computerized training of working memory in children with ADHD-A randomized, controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry, 10*, 177–186.

Kolahi, J., Bang, H., & Park, J. (2009). Towards a proposal for assessment of blinding success in clinical trials: Up-to-date review. *Community Dentistry and Oral Epidemiology, 37*(6), 477–484. https://doi.org/10.1111/j.1600-0528.2009.00494.x

Leibowitz, K. A., Hardebeck, E. J., Goyer, J. P., & Crum, A. J. (2019). The role of patient beliefs in open-label placebo effects. *Health Psychology, 38*(7), 613–622. https://doi.org/10.1037/hea0000751

Lilienthal, L., Tamez, E., Shelton, J. T., Myerson, J., & Hale, S. (2013). Dual n-back training increases the capacity of the focus of attention. *Psychonomic Bulletin & Review, 20*(1), 135–141. https://doi.org/10.3758/s13423-012-0335-6

Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of "far transfer": Evidence from a meta-analytic review. *Perspectives on Psychological Science, 11*(4), 512–534. https://doi.org/10.1177/1745691616635612

Miller, H. (2011). Self-control and health outcomes in a nationally representative sample. *American Journal of Health Behavior, 35*(1), 15–27. https://doi.org/10.5993/AJHB.35.1.2

Morris, D., Fraser, S., & Wormald, R. (2007). Masking is better than blinding. *BMJ, 334*(7597), 799–799.https://doi.org/10.1136/bmj.39175.503299.94

Nichols, A. L., & Maner, J. K. (2008). The good-subject effect: Investigating participant demand characteristics. *The Journal of General Psychology, 135*(2), 151–165.

Novak, E., & Tassell, J. (2015). Using video game play to improve education-majors' mathematical performance: An experimental study. *Computers in Human Behavior, 53*, 124–130. https://doi.org/10.1016/j.chb.2015.07.001

Oei, A. C., & Patterson, M. D. (2013). Enhancing cognition with video games: A multiple game training study. *PLoS ONE, 8*(3), e58546. https://doi.org/10.1371/journal.pone.0058546

Oei, A. C., & Patterson, M. D. (2015). Enhancing perceptual and attentional skills requires common demands between the action video games and transfer tasks. *Frontiers in Psychology, 6*, 113. https://doi.org/10.3389/fpsyg.2015.00113

Parong, J., Seitz, A. R., Jaeggi, S. M., & Green, C. S. (2022). *Expectation Effects in Working Memory Training, 119*(37), 10.

Parong, J., Vodyanyk, M., Green, C. S., Jaeggi, S. M., & Seitz, A. R. (2023). Experimenter effects. In A. L. Nichols & J. Edlund (Eds.), *The Cambridge Handbook of Research Methods and Statistics for the Social and Behavioral Sciences* (1st ed., pp. 224–243). Cambridge University Press. https://doi.org/10.1017/9781009010054.012

Rabipour, S., & Davidson, P. S. R. (2015). Do you believe in brain training? A questionnaire about expectations of computerised cognitive training. *Behavioural Brain Research, 295*, 64–70. https://doi.org/10.1016/j.bbr.2015.01.002

Ramos-Galarza, C., Acosta-Rodas, P., Bolaños-Pasquel, M., & Lepe-Martínez, N. (2019). The role of executive functions in academic performance and behaviour of university students. *Journal of Applied Research in Higher Education, 12*(3), 444–455. https://doi.org/10.1108/JARHE-10-2018-0221

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., Kane, M. J., & Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General, 142*(2), 359–379. https://doi.org/10.1037/a0029082

Riggs, N. R., Spruijt-Metz, D., Sakuma, K.-L., Chou, C.-P., & Pentz, M. A. (2010). Executive cognitive function and food intake in children. *Journal of Nutrition Education and Behavior, 42*(6), 398–403. https://doi.org/10.1016/j.jneb.2009.11.003

Rohde, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability.*Intelligence, 35*(1), 83–92. https://doi.org/10.1016/j.intell.2006.05.004

Sala, G., Tatlidil, K. S., & Gobet, F. (2018). Video game training does not enhance cognitive ability: A comprehensive meta-analytic investigation. *Psychological Bulletin, 144*(2), 111–139. https://doi.org/10.1037/bul0000139

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2019). Training working memory for 100 days: The COGITO study. In J. M. Novick, M. F. Bunting, M. R. Dougherty, & R. W. Engle (Eds.), *Cognitive and Working Memory Training* (1st ed., pp. 40–57). Oxford University PressNew York. https://doi.org/10.1093/oso/9780199974467.003.0003

Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods, 22*(2), 322–339. https://doi.org/10.1037/met0000061

Schwaighofer, M., Fischer, F., & Bühner, M. (2015). Does working memory training transfer? A meta-analysis including training conditions as moderators. *Educational Psychologist, 50*(2), 138–166. https://doi.org/10.1080/00461520.2015.1036274

Sella, F., Gal, R., & Roi, C. K. (2021). When randomisation is not good enough: Matching groups in intervention studies. *Psychonomic Bulletin & Review, 9*, 2085–2093. https://doi.org/10.3758/s13423-021-01970-5

Shute, V. J., Ventura, M., & Ke, F. (2015). The power of play: The effects of portal 2 and lumosity on cognitive and noncognitive skills. *Computers & Education, 80*, 58–67. https://doi.org/10.1016/j.compedu.2014.08.013

Spence, I., & Feng, J. (2010). Video games and spatial cognition. *Review of General Psychology, 14*(2), 92–104. https://doi.org/10.1037/a0019491

Stepankova, H., Lukavsky, J., Buschkuehl, M., Kopecek, M., Ripova, D., & Jaeggi, S. M. (2014). The malleability of working memory and visuospatial skills: A randomized controlled study in older adults. *Developmental Psychology, 50*(4), 1049–1059. https://doi.org/10.1037/a0034913

Stordal, K. I., Lundervold, A. J., Egeland, J., Mykletun, A., Asbjørnsen, A., Landrø, N. I., Roness, A., Rund, B. R., Sundet, K., Oedegaard, K. J., & Lund, A. (2004). Impairment across executive functions in recurrent major depression. *Nordic Journal of Psychiatry, 58*(1), 41–47. https://doi.org/10.1080/08039480310000789

Strobach, T., Frensch, P. A., & Schubert, T. (2012). Video game practice optimizes executive control skills in dual-task and task switching situations. *Acta Psychologica, 140*(1), 13–24. https://doi.org/10.1016/j.actpsy.2012.02.001

Taylor Tavares, J. V., Clark, L., Cannon, D. M., Erickson, K., Drevets, W. C., & Sahakian, B. J. (2007). Distinct profiles of neurocognitive function in unmedicated unipolar depression and bipolar II depression. *Biological Psychiatry, 62*(8), 917–924. https://doi.org/10.1016/j.biopsych.2007.05.034

Thorell, L. B., Lindqvist, S., Bergman Nutley, S., Bohlin, G., & Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental Science, 12*(1), 106–113. https://doi.org/10.1111/j.1467-7687.2008.00745.x

Tiraboschi, G. A., Fukusima, S. S., & West, G. L. (2019). An expectancy effect causes improved visual attention performance after video game playing. *Journal of Cognitive Enhancement, 3*(4), 436–444. https://doi.org/10.1007/s41465-019-00130-x

Titz, C., & Karbach, J. (2014). Working memory and executive functions: Effects of training on academic achievement. *Psychological Research, 78*(6), 852–868. https://doi.org/10.1007/s00426-013-0537-1

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. https://doi.org/10.1037/h0101826

Unsworth, N., Redick, T. S., McMillan, B. D., Hambrick, D. Z., Kane, M. J., & Engle, R. W. (2015). Is playing video games related to cognitive abilities? *Psychological Science, 26*(6), 759–774. https://doi.org/10.1177/0956797615570367

Vodyanyk, M., Cochrane, A., Corriveau, A., Demko, Z., & Green, C. S. (2021). No evidence for expectation effects in cognitive training tasks. *Journal of Cognitive Enhancement, 5*(3), 296–310. https://doi.org/10.1007/s41465-021-00207-6

Watanabe, T., & Sasaki, Y. (2015). Perceptual learning: Toward a comprehensive theory. *Annual Review of Psychology, 66*(1), 197–221. https://doi.org/10.1146/annurev-psych-010814-015214

Weiss, R. (1997). Bayesian sample size calculations for hypothesis testing. *Journal of the Royal Statistical Society: Series D (The Statistician), 46*(2), 185–191. https://doi.org/10.1111/1467-9884.00075

Wu, S., Cheng, C. K., Feng, J., D'Angelo, L., Alain, C., & Spence, I. (2012). Playing a first-person shooter video game induces neuroplastic change. *Journal of Cognitive Neuroscience, 24*(6), 1286–1293. https://doi.org/10.1162/jocn_a_00192

Zhang, R.-Y., Chopin, A., Shibata, K., Lu, Z.-L., Jaeggi, S. M., Buschkuehl, M., Green, C. S., & Bavelier, D. (2021). Action video game play facilitates "learning to learn." *Communications Biology, 4*(1), 1154. https://doi.org/10.1038/s42003-021-02652-7

Zunhammer, M., Ploner, M., Engelbrecht, C., Bock, J., Kessner, S. S., & Bingel, U. (2017). The effects of treatment failure generalize across different routes of drug administration. *Science Translational Medicine, 9*(393), eaal2999. https://doi.org/10.1126/scitranslmed.aal2999