bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Few Effects of a 5-week Computerized Cognitive Training Program in Healthy Older Adults

Sheida Rabipour[1], Cassandra Morrison[1,2], Jessica Crompton[1], Marcelo Petrucelli[3], Murillo de Oliveira Gonçalves Germano[4], Anita Popescu[3], and Patrick S. R. Davidson[1,2]

[1]School of Psychology, University of Ottawa, Ottawa, Canada

[2]Bruyère Research Institute, Bruyère Continuing Care, Ottawa, Canada

[3]Department of Computer Science, University of Ottawa, Ottawa, Canada

[4]Department of Mechanical Engineering, University of Ottawa, Ottawa, Canada

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Computerized cognitive training programs are becoming increasingly popular and practical for cognitive aging. Nevertheless, basic questions remain about the benefits of such programs, and about the degree to which participant expectations might influence training and transfer. Here we examined a commercial cognitive training program (*Activate*) in a 5-week double-blind, pseudo-randomized placebo-controlled trial. Based on a priori power analysis, we recruited 99 healthy older adults 59-91 years of age (M = 68.87, SD = 6.31; 69 women), assigning them to either the intervention or an active control program (Sudoku and n-back working memory exercises). We subdivided both groups into high and low expectation priming conditions, to probe for effects of participants' expectations on training and transfer. We assessed transfer using a battery of standard neuropsychological and psychosocial measures that had been agreed to by the training program developers. We planned and pre-registered our analyses (on osf.io). The majority (88%) of participants progressed through the training, and most provided positive feedback about it. Similarly, the majority (80%) of participants believed they were truly training their brains. Yet, transfer of training was minimal. Also minimal were any effects of expectations on training and transfer, although participants who received high expectation priming tended to engage more with their assigned program overall. Our findings suggest limited benefits of *Activate* training on cognition and psychosocial wellbeing in healthy older adults, at least under the conditions we used.

Keywords

Aging; Brain training; Computerized cognitive training; Executive function; Expectation; Placebo effect; Working memory.

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Subtle decline in several cognitive domains is typical even in healthy aging (Anderson & Craik, 2017; Davidson & Winocur, 2017; Drag & Bieliauskas, 2010; Toepper, 2017). To prevent or slow this cognitive decline, computerized interventions are increasingly becoming popular among researchers, clinicians, and older adults themselves. Such Computerized Cognitive Training (CCT) represents an attractive alternative to medication and other interventions that may be costly and have serious side effects (National Academies of Sciences, 2017). It is therefore unsurprising that older adults are prominent targets of – and subscribers to – commercialized CCT programs (e.g., Lumosity, Cognifit, Brain Age; Shah, Weinborn, Verdile, Sohrabi, & Martins, 2017).

Nevertheless, basic questions remain about CCT. An obvious one is whether such programs actually work. When people ask this question, they are usually asking about transfer of training: That is, will training lead to improvement on tasks that were not trained, but that rely on the same underlying cognitive processes as those that underwent training? Answering this question has been difficult due to the poor methodological quality of previous research on CCT. For example, studies have often suffered from insufficient control conditions – notably, small sample sizes, frequent lack of an active control group, and inadequate consideration of potential placebo effects (W. R. Boot, Simons, Stothart, & Stutts, 2013; Green et al., 2019; Motter, Devanand, Doraiswamy, & Sneed, 2016; National Academies of Sciences, 2017; Simons et al., 2016). Importantly, cognitive training trials have frequently employed passive (e.g., waitlist) controls as a comparison for the training group(s). This practice is problematic as it permits only an evaluation of test-retest effects without providing any insight into nonspecific factors related to participating in a program.

Here we report a trial of the commercial CCT program *Activate* in older adults, in which we addressed many of these limitations, including the largely ignored potential confound of

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

participant expectations. In particular, we aimed to evaluate the real-world practicality of such training by examining transfer effects, beyond expected practice (i.e., test-retest) effects (Collie, Maruff, Darby, & McStephen, 2003), using standard neuropsychological measures.

Activate Program

Although many commercial CCT programs are available (for a partial list and review, see Simons et al., 2016), we selected the *Activate* web-based program because of its focus on training working memory (WM), sustained and divided attention, inhibition, and visuospatial perception. Many of these functions decline even in healthy aging (Davidson & Winocur, 2017; Reuter-Lorenz, Festini, & Jantz, 2016; Zanto, Chadick, & Gazzaley, 2014). Specifically, the *Activate* games are divided into four categories, targeting: i) spatial WM; ii) visual pursuit and category sorting; iii) visual pursuit and inhibition; and iv) "pattern completion" (i.e., determining the image that best fits into an empty space, to complete a pattern).

In our view, the *Activate* program had two advantages: First, the training is adaptive. In other words, the program adjusts for individual performance on each task, remaining at a certain level of difficulty as long as participant accuracy and response times continue to improve at a certain rate. The program only increases level of difficulty once performance has reached a plateau, rather than advancing to the next level after a predetermined number of correct trials. This approach allows participants to master each task at a given level of difficulty before moving on to subsequent levels. Researchers have argued that such adaptive training methods are particularly likely to produce transfer of training (Heinzel et al., 2014; Lovden, Backman, Lindenberger, Schaefer, & Schmiedek, 2010; Morrison & Chein, 2011). Second, the *Activate* program had already shown promise of transfer in older adults with depression, with training reported to reduce depressive symptoms and measurably improve cognitive control (as indexed by the Trail-Making and Stroop

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

tests; Morimoto et al., 2016; Morimoto et al., 2014). A recent report further demonstrated the potential of *Activate* to improve children's performance on executive function measures in the classroom setting (Kavanaugh, Tuncer, & Wexler, 2018). Despite these promising preliminary reports, no other evaluations of this program currently exist (Simons et al., 2016). Thus, we sought to replicate these previous reports by including the neuropsychological tasks from those studies (i.e., Trail-Making and Stroop tests), and came to a consensus with the *Activate* developers on other tasks that should show transfer in older adults.

Expectations & Placebo Effects

Much of the CCT literature has been criticized for having relatively low methodological quality (e.g., Motter et al., 2016; Simons et al., 2016). Perhaps the most serious problems revolve around insufficient control conditions, including lack of an appropriate active control group to distinguish intervention effects from those of placebo or other non-specific factors, and failure to account for participants' expectations of training outcomes (W. R. Boot et al., 2013; Foroughi, Monfort, Paczynski, McKnight, & Greenwood, 2016). Previous surveys have suggested that older adults have optimistic expectations of CCT (Rabipour, Andringa, Boot, & Davidson, 2017; Rabipour & Davidson, 2015). Thus, what appear to be transfer effects following CCT could, instead, merely represent placebo effects. To our knowledge, expectations have rarely – if ever – been explicitly addressed in CCT trials (for the few existing examples, see W. R. Boot et al., 2013; Foroughi et al., 2016). A powerful way to accomplish this is with a balanced-placebo design (Rohsenow & Marlatt, 1981), using a randomized double-blind placebo-controlled design and sub-dividing both the treatment and placebo control groups such that half of participants in each group are told they are in the treatment condition and the other half told they are in the placebo control group. Such a design helps one tease apart the true effects of the intervention from those of expectations.

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The Present Study

Here we examined a 5-week CCT program in healthy older adults actively seeking preventative cognitive interventions. Using a balanced-placebo design, we aimed to elucidate the potential influence of such expectations on cognitive performance following participation in CCT. We took particular care in our study design and data collection to ensure: i) reasonable statistical power; ii) use of a dynamic CCT program adaptive to participant performance; iii) use of an active (rather than less-rigorous passive) control condition (Redick, 2015; Shipstead, Redick, & Engle, 2010), comprised of a computerized program that would believably pass as cognitive training; iv) control of potential expectancy-mediated placebo effects (W. R. Boot et al., 2013; Green et al., 2019); v) use of consensus transfer measures that the developers of the CCT program agreed should reflect the cognitive processes being trained by their exercises, and that are standard measures used in both clinical and experimental studies, including in previous evaluations of the *Activate* program; vi) replication and extension of previous findings on *Activate* improving executive function in older adults (Morimoto et al., 2014), and of our own work on older adults having higher expectations of CCT (Rabipour & Davidson, 2015); and vii) double-blind design. These steps follow best practice recommendations (e.g., Green et al., 2019; National Academies of Sciences, 2017).

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

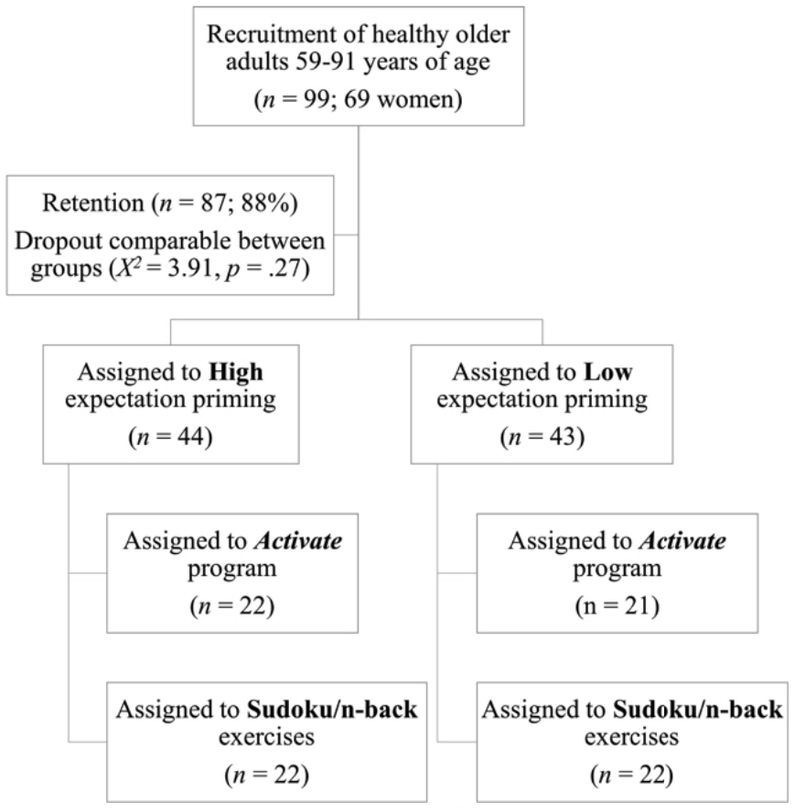

Figure 1. Participant recruitment and group assignment with final sample.

Methods

Participants

We recruited 99 healthy older adults (M age=68.88, SD=6.31 years; 69 women) with normal or corrected-to-normal vision and hearing from the Ottawa community via ads and flyers (Figure 1). We screened all participants based on the following exclusion criteria: i) diagnosis of memory impairment or cognitive decline; ii) diagnosis of neurological disorder; iii) experience of stroke or transient ischemic attack; and iv) experience of injury or operation resulting in long-term changes in memory or cognitive function.

A priori power analyses using G*Power version 3.1 determined that a minimal sample of 84 participants (i.e., 21 older adults per group) would be required for repeated-measures ANOVA with within-between subjects interactions, based on a moderate effect size (Cohen's f(V)=.45) at

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

80% power and alpha=.01 to account for multiple measurements. This interaction between assigned group (see below) x time (pre vs. post training) – and, in particular, between time and training program – would provide the strongest evidence of *Activate* having specific transfer effect(s).

Experimental Design and Protocol

The study was a double blind randomized controlled trial (registered at clinicaltrials.gov, ID: NCT0220571), crossing expectation priming (high vs. low) with CCT program (*Activate* vs. placebo-controlled Sudoku and n-back exercises). After providing consent, participants were assigned to one of two expectation priming conditions: i) High expectation priming, in which participants were told they would receive a type of CCT known to improve performance; and ii) Low expectation priming, in which participants were told they would receive a type of CCT with no known benefits. We then assigned participants to receive one of two programs: the commercialized *Activate* CCT intervention, or an active control (see below).

We followed a pseudo-randomization procedure, aiming to balance the age and sex of participants across groups as best as possible. Importantly, our expectation priming conditions and comparison of two separate CCT programs enabled blinding of participants with respect to their experimental condition. To avoid divulging the nature of the study and potentially skewing the results, participants who entered the study with a friend or family member, or who were referred to the study by a past participant, were assigned to the same group. To maintain blinding for researchers, we ensured that those examining participants would not be present during the training sessions and would have no contact with participants outside the pre-and post-training assessments.

Both CCT programs comprised five 40-minute training session per week, over five weeks. The 25 sessions each involved two 20-minute periods of training, separated by a brief break (i.e.,

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

several minutes) if desired. Each week, participants were instructed to complete a minimum of four home sessions in addition to one lab session, carried out in a small group setting at the University of Ottawa main campus (i.e., where several participants would complete their training together in the same room, on individual computers). The rationale for the weekly lab sessions twofold: First, the supervised lab sessions helped us ensure that participants were adhering to the correct training protocol and were not experiencing any issues related to the program or understanding of the training tasks. Second, previous research has suggested that older adults benefit from CCT to a greater extent when completing the program, at least in part, in supervised group sessions conducted at a center (Lampit, Hallock, & Valenzuela, 2014; Wadley et al., Dividing the home and lab sessions in this manner permitted us to balance between convenience and quality control. Similarly, we selected the 5-week time frame to promote practicality and minimize participant attrition, while permitting a reasonable amount of training (i.e., 1000min or ±17h), comparable to other training studies and plausibly leading to transfer (e.g., Kavanaugh et al., 2018).

We administered behavioural assessments at baseline and following completion of the CCT program. We also performed a partial long-term follow-up assessment, wherein we asked participants to complete select questionnaires six months following the completion of the program. At the time of this long-term follow-up, participants were contacted via telephone to provide final comments on their experience and perspective on the effects of the program via an informal interview.

In order to maximize intrinsic motivation and avoid possible performance decrements associated with receipt of extrinsic rewards (Jaeggi, Buschkuehl, Shah, & Jonides, 2014), participants received no monetary or other compensation for participation in the CCT program. However, participants were able to retain free access to the program beyond their participation in the study. Moreover, to compensate for time spent in the research lab at the University of

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Ottawa, participants were offered free parking as well as a chance to enter a lottery to win a 50$CAD gift card. We received ethical approval to conduct this study from the Research Ethics Boards at the University of Ottawa and the Bruyère Research Institute.

Expectation Assessment and Priming

Participants rated their expectations of CCT effectiveness using the Expectation Assessment Scale (EAS), a validated questionnaire (Rabipour, Davidson, & Kristjansson, 2018) we have previously used to measure and prime expectations with written messages implying that the intervention is either highly effective or not expected to produce any meaningful change (see Rabipour, Andringa, et al., 2017; Rabipour & Davidson, 2015). We administered the EAS on three occasions: i) at baseline; ii) after participants received the expectation priming messages (High or Low); and iii) after completing the 5-week CCT program. Participants rated their expectations of outcomes on a scale from 1-7 (1 = "completely unsuccessful," 2 = "fairly unsuccessful," 3 = "somewhat unsuccessful," 4 = neutral/"I have absolutely no expectations," 5 = "somewhat successful," 6 = "fairly successful," 7 = "completely successful"). We probed expectations on seven cognitive domains: i) general cognitive function; ii) memory; iii) concentration; iv) distractibility; v) reasoning ability; vi) multitasking ability; and vii) performance in everyday activities. Following the training program, we probed participants about their experience using a feedback questionnaire (see supplemental files).

Training Programs

Activate training

The intervention, the pirate-themed *Activate* web program (C8 Sciences), comprised six adaptive computer games involving WM, sustained and divided attention, inhibition, and visuospatial

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

perception. Specifically, the games are divided into four categories based on targeted cognitive functions: i) spatial WM (two games); ii) visual pursuit and category sorting (one game); iii) visual pursuit and inhibition (two games); and iv) pattern completion (one game).

The *Activate* intervention aims to account for individual differences in performance on each task: each task remains at a level of difficulty as long as participant accuracy and response times continue to improve, rather than advance after a predetermined number of correct trials. The program increases level of difficulty only once performance reaches a plateau. Originally created to enhance school performance in children with impulse-control impairments, the program has more recently been applied to older adults and appears to reduce depressive symptoms as well as executive dysfunction in patients with geriatric depression (Morimoto et al., 2016; Morimoto et al., 2014).

During each session, participants were presented with a choice of three possible games. The selected game was played for five minutes, after which the program terminated the game and presented participants with a new choice of three games. Once participants completed four 5-minute game blocks (i.e., 20 minutes of training), the program automatically logged participants out. We asked participants to complete two 20-minute training periods in immediate succession (with the possibility of a brief break in between), at least five different days per week, resulting in five weekly 40-minute sessions in total. Although participants were able to select which set of games to play throughout each session, the program limited choice if participants strongly favoured certain games over others, to ensure roughly equivalent amount of training with all games.

Sudoku and single n-back training

The active control program comprised non-adaptive computer games with no unifying theme.

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Participants had a choice of two game types: Sudoku or single n-back WM exercises. We selected these games based on popular conceptions that they broadly improve cognitive functions (Walter R. Boot et al., 2016), coupled with little empirical evidence supporting this notion (Shipstead, Redick, & Engle, 2012; Souders et al., 2017). Unlike the intervention groups, participants assigned to the active control condition selected games entirely at their own discretion for each session. We nevertheless encouraged participants to split their sessions evenly between Sudoku puzzles and n-back games. Similar to the *Activate* training, participants were automatically logged out of the game portal after 20 minutes of task engagement, and were asked to complete two 20-minute training blocks in immediate succession (again, with possibility of a brief break), at least five different days per week.

For the Sudoku puzzles, participants had the option to select “easy”, “medium”, “hard”, or “evil” levels, as per traditional Sudoku. While we wanted to ensure adequate engagement with the Sudoku puzzles, we also sought to further reduce the possibility that this control program would enhance cognitive function or benefit participant performance on untrained tasks. Thus, based on pilot testing, the Sudoku puzzles included in all levels beyond “easy” were, in reality, randomized to generate an easy-level puzzle on 80% of trials, and a puzzle of the selected level on the remaining 20% of trials. Moreover, the “evil” level was, in reality, equivalent to “hard” (i.e., we did not actually include any “evil” level Sudoku puzzles in our program). Notably, because the control program was intended solely as a comparison rather than training activity, we prioritized minimizing difficulty while maintaining engagement. Our analyses of these data reflect the level of Sudoku participants actually completed, rather than the level they selected.

For the n-back tasks, participants had the option of selecting 2-, 3-, or 4-back tasks. Meta-analyses suggest minimal evidence of transfer, especially for single-domain n-back (Redick & Lindsey, 2013), whereas our goal was to evaluate far transfer effects. The type of n-back was randomized to comprise: i) a randomly generated string of letters and numbers; ii) a

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

randomly generated string of pictures; or iii) a spatial n-back task involving a cartoon brain image randomly moving across a 3x3 grid. Due to technical issues with our database, we were unable to retain data regarding level of n-back for the majority of our participants; we analyzed these data based on n-back type.

Transfer Measures

We included screening measures of global cognition and language function at baseline as potential discriminating variables in our subsequent analyses. We assessed global cognition using the Montreal Cognitive Assessment (MoCA), a brief cognitive screening tool sensitive to mild declines in cognitive function (Nasreddine et al., 2004). Based on advice from our colleagues, we also used the 15-item Boston Naming Test (Lansing, Ivnik, Cullum, & Randolph, 1999) as a screen for language function in our final 53 participants.

As indicators of transfer, we assessed performance on commonly used neuropsychological tests and subjective reports of cognition, quality of life (QOL), and affect. Here we focus on the neuropsychological tests (for the other data, see supplemental files). These measures were agreed on *a priori* by the lead authors (S.R. and P.S.R.D) and the creators of the *Activate* program as reasonable indicators of transfer:

Wisconsin Card Sorting Task (WCST). The WCST, which requires participants to categorize cards by colour, shape, or number using only computer feedback, is a measure of executive function and cognitive flexibility (Rhodes, 2004). We evaluated number of categories completed, an index sensitive to age-related changes in executive function (Rhodes, 2004).

California Verbal Learning Test, second edition (CVLT-II). The CVLT assesses verbal memory by presenting the participant with two lists of 16 words, divided into four semantic categories, over several trials. The full version of the CVLT, used in the present study, involves a short-and long-delay component separated by 15-25 minutes. Scores are typically lower with

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

increasing age and lower level of education, as well as in men compared to women (Lamar, Resnick, & Zonderman, 2003). We selected to evaluate the semantic clustering (Stricker, Brown, Wixted, Baldo, & Delis, 2002) and long-delay cued recall sub-components of the CVLT, which index strategy use (i.e., organizational clustering) and recall ability, respectively.

Digit Span Task. The digit span task, a subtest of the Wechsler Adult Intelligence Scale, is an index of WM capacity. For the forward digit span (FDS) task, a measure of short-term recall, participants repeat a string of numbers that the examiner reads aloud; in the backward digit span (BDS) task, an index of WM capacity, participants recite the string of numbers in the reverse order. In the present study, participants completed both the FDS and BDS tasks, but only scores for the BDS were included in our analyses as a measure of executive function (Glisky, Polster, & Routhieaux, 1995).

Trail Making Test (TMT). The TMT requires participants to draw a straight line between numbers (Trail A) and alternating numbers and letters (Trail B) in sequential order, as quickly as possible. The time taken to complete each trail provides a measure of visuospatial capacity, executive control, and cognitive flexibility (Kortte, Horner, & Windham, 2002). Here we administered both Trails A and B, but included only Trail B time for our final analyses, as per previous studies of *Activate* (Morimoto et al., 2016; Morimoto et al., 2014).

Controlled Oral Word Association Test (COWAT) & Animal Naming Test. The COWAT measures letter fluency (LF) with the generation of words that begin with a given letter within 60 seconds. Participants name as many words as possible that begin with the letter “F”, then “A”, and finally “S”. The animal-naming test, administered immediately after the COWAT, is a measure of semantic fluency that requires participants to name as many animals as possible in 60 seconds. Scores represent the total number of admissible words reported and appear sensitive to age-related changes in verbal fluency. Although participants completed both the LF and animal naming tasks, only scores for the LF (analyzed as an average number of words per letter) were

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

included in our analyses as an index of executive function (Glisky et al., 1995; McCabe, Roediger, McDaniel, Balota, & Hambrick, 2010).

Stroop Interference Task. One of the most commonly used measures in cognitive psychology, the Stroop task (Stroop, 1935) measures attention, processing speed, inhibition, and executive control. The test proceeds in three timed parts of 45 seconds each: In the first part, which measures reading speed, the participant is given a list of colour words and must read as many words as possible within the allotted time. In the second part, which measures colour-naming speed, the participant must name the ink color of printed "X"s as quickly as possible. In the final part – the interference condition – the participant must quickly name the color of the ink in which the colour words are written, without reading the words themselves (e.g., say "blue" when seeing the word "green" written in blue ink). The version used in the present study contains 120 words per part; scores in each part of the test are combined to create an interference score. Because it is brief, requires very little education, and is not culturally biased, this test is a useful screen for neuropsychological deficits. Older adults appear to have slower interference scores, which may reflect decreased efficiency of inhibitory processing (Spieler, Balota, & Faust, 1996).

Letter-Number Sequencing. In a subset of our participants (n=69), we additionally explored performance on letter-number sequencing, a subtest of the Wechsler Adult Intelligence Scale. This brief, standardized measure of executive function indexes verbal and visuospatial WM performance, as well as processing speed (Crowe, 2000). During the task, examiners read a series of scrambled letters and numbers to participants, who must then recite the numbers in sequential order, followed by the letters in alphabetical order.

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Statistical Analysis

We performed analyses using IBM SPSS Statistics, Inc. version 24, R version 3.1.1, and JASP version 0.9.2. We pre-registered our analysis plan on Open Science Framework, as part of the "OSF Pre-registration Challenge".

We conducted repeated-measures multivariate ANOVA (MANOVA) on each category of outcomes (i.e., expectation ratings, training tasks, and neuropsychological and psychosocial transfer measures) with expectation priming and CCT conditions as between-subjects factors, and time as a within-subjects factor. We followed up significant results with simple effects ANOVA and t-tests between the groups of interest, and non-significant results with Bayesian repeated measures ANOVA using the default parameters in JASP. Where applicable, we used the Holm-Bonferroni correction for multiple comparisons and Greenhouse-Geisser correction for sphericity.

Results

MANOVA between assigned program (*Activate* vs. control) and expectation priming condition (high vs. low) indicated no significant group differences in age, years of education, cognitive function (MoCA and Boston Naming scores), or medication use (Wilk's $\lambda \geq .867$, $F_{(5,35)} \leq 1.08$, $p \geq .39$, $BF_{10} \leq .479$; Table 1). Chi-squared analysis similarly revealed no significant group differences in sex ($X^2 = 1.06$, p=.787). Self-reports of medication use indicated that the majority of participants (78/87=90%) did not take anticholinergic medication. Moreover, the distribution of Anticholinergic Cognitive Burden scores (Campbell, Maidment, Fox, Khan, & Boustani, 2013) did not significantly differ between groups ($X^2 = 2.31$, p=.511).

Participant retention rate was high: after dropout due to scheduling conflicts (n=5), health

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

issues (n=4), loss of interest (n=2), and other personal emergencies (n=1), 88% (87/99; 60 women) of our sample completed the study and were included in our final analyses.

Table 1. Participant demographics. Groups did not significantly differ with respect to distribution of sex, mean age, years of education, global cognition and memory, language function, and self-reported number of medications ingested.

Group	**Program**	**Expectation Priming**	***n* (women)**	**Age**	**Years of Education**	**MoCA Global Score**	**MoCA Memory Score**	**Boston Naming Test**	**Medication Number**
1	*Activate*	High	22 (17)	67.59 (6.89)	16.55 (2.48)	27.5 (2.69)	3.23 (1.82)	14.55 (.52)	2.11 (1.84)
2	*Activate*	Low	22 (14)	69.59 (6.91)	17.64 (3.53)	26.86 (2.77)	2.45 (2.02)	13.82 (.98)	3.00 (3.25)
3	Sudoku/n-back	High	21 (14)	68.95 (4.44)	15.86 (2.97)	27.29 (2.45)	3.24 (1.79)	14.45 (.82)	2.29 (2.37)
4	Sudoku/n-back	Low	22 (15)	69.64 (7.22)	16.77 (2.41)	26.48 (3.19)	3.00 (1.90)	14.18 (.98)	2.41 (2.50)

Expectations of Outcomes

Participants in all groups were largely optimistic of CCT outcomes at baseline (Figure 2), with ratings in all cognitive domains significantly above neutral for participants in all groups ($t_{21}\geq 2.11$, $p\leq .047$, Cohen's $d\geq .45$), with the exception of ratings for "multitasking" in participants assigned to *Activate* training and low expectation priming ($t_{21}=.84$, $p=.41$, $BF_{10}=.031$). Mean expectation ratings did not significantly differ across groups at baseline.

We performed repeated measures MANOVA comparing expectation ratings across the seven cognitive domains between expectation priming condition and CCT program at three times: baseline, after receiving the expectation priming message, and after completing the CCT. This analysis revealed a significant main effect of time (Wilk's $\lambda=.349$, $F_{(14,70)}=9.31$, $p<.0001$, $\eta_p^2=.65$) as well as a significant interaction between time and expectation condition (Wilk's $\lambda=.618$, $F_{(14,70)}=3.08$, $p=.001$, $\eta_p^2=.38$; Figure 2). We did not find a main effect of program or of expectation condition, or an interaction effect, on expectation ratings in any cognitive domain,

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

after correcting for multiple comparisons ($BF_{10} \leq .906$).

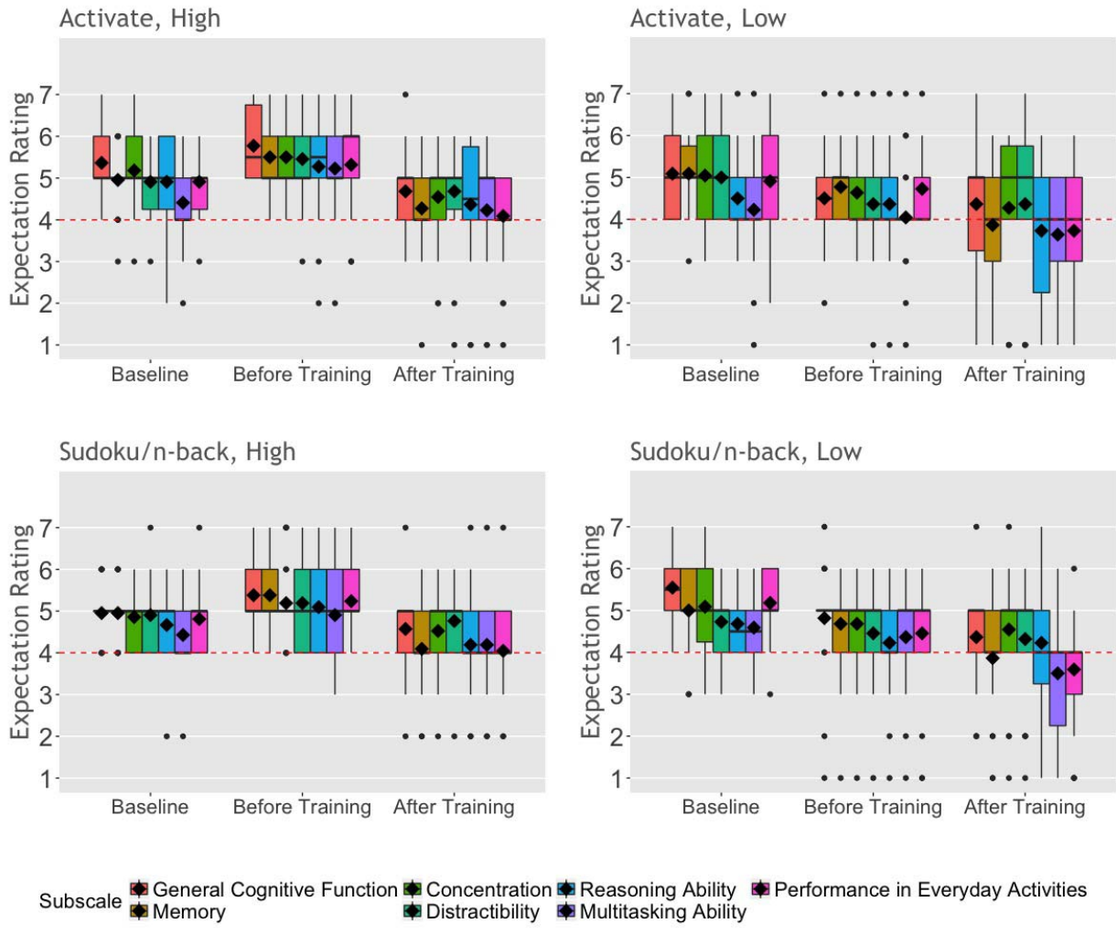

Figure 2. Participant EAS ratings at baseline, after receiving expectation priming (i.e., before training), and following completion of their assigned 5-week CCT program. Ratings of expected outcomes were on a scale of "1" ("completely unsuccessful") to "7" ("completely successful") for each subscale (i.e., cognitive domain). Red dashed lines represent a neutral rating of "4" ("no expectation"). Solid lines represent group medians; diamonds represent group means.

Probing the results for each cognitive domain, we found a significant effect of time ($F_{(1.58,131.4)} \geq 11.17$, $p < .0001$, $\eta_p^2 \geq .104$) and significant interaction between time and expectation condition ($F_{(1.58,152.15)} \geq 3.50$, $p \leq .044$, $\eta_p^2 \geq .04$) for all cognitive domains, except for "distractibility", where the effect of time ($F_{(1.58,130.7)} = 3.30$, $p = .051$, $\eta_p^2 = .038$; $BF_{10} = .786$) was marginal.

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Pairwise comparisons showed that participants in the high expectation condition significantly increased their expectation ratings immediately after receiving the priming message, compared to baseline (i.e., from baseline to before training) in all cognitive domains ($M_{diff} \geq .33$; $p \leq .003$). Conversely, after correcting for multiple comparisons, participants who received low expectation priming decreased their expectation ratings only for the "general cognitive function" domain ($M_{diff} \geq .66$; $p \leq .0001$), after receiving the low expectation priming message. After correcting for multiple comparisons, we did not find group differences in ratings for any cognitive domain following completion of the program.

Training Performance

Engagement with Training Tasks

Using repeated measures MANOVA, we evaluated the amount of time participants in each expectation priming condition engaged in the *Activate* games over the five weeks (Figure 3a). We found a main effect of week (Wilk's λ=.309, $F_{(16,27)}$=3.77, p=.001, η_p^2=.69) and of expectation priming (Wilk's λ=.762, $F_{(4,39)}$=3.04, p=.028, η_p^2=.24), but no interaction between expectation and week ($F_{(16,27)}$=1.63, p=.128, BF_{10}=.051). Further examination revealed the effect of week to be significant for the "visual pursuit / category sorting" ($F_{(4,168)}$=2.93, p=.022, η_p^2=.065) and "visual pursuit / inhibition" ($F_{(4,168)}$=13.81, p<.0001, η_p^2=.25) categories, and the effect of expectation to be significant for the "visual pursuit / category sorting" games ($F_{(1,42)}$=5.76, p=.021, η_p^2=.12). Specifically, participants who received high expectation priming engaged significantly more with the "visual pursuit / category sorting" games, compared to those in the low expectation priming condition. Regarding the effect of week, participants engaged in

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the "visual pursuit / category sorting" games significantly more during the fourth week of the program, compared to the first, second, and fifth weeks ($t_{43} \geq 2.27$, $p \leq 0.028$, Cohen's $d \geq .69$); for the "visual pursuit / inhibition" category, participants engaged in the games significantly less in the fourth and fifth weeks ($t_{43} \geq 2.67$, $p \leq 0.011$, Cohen's $d \geq .81$), compared to the first three. Based on these results, we included amount of time played per week as a covariate in analyses of performance outcomes (see below).

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

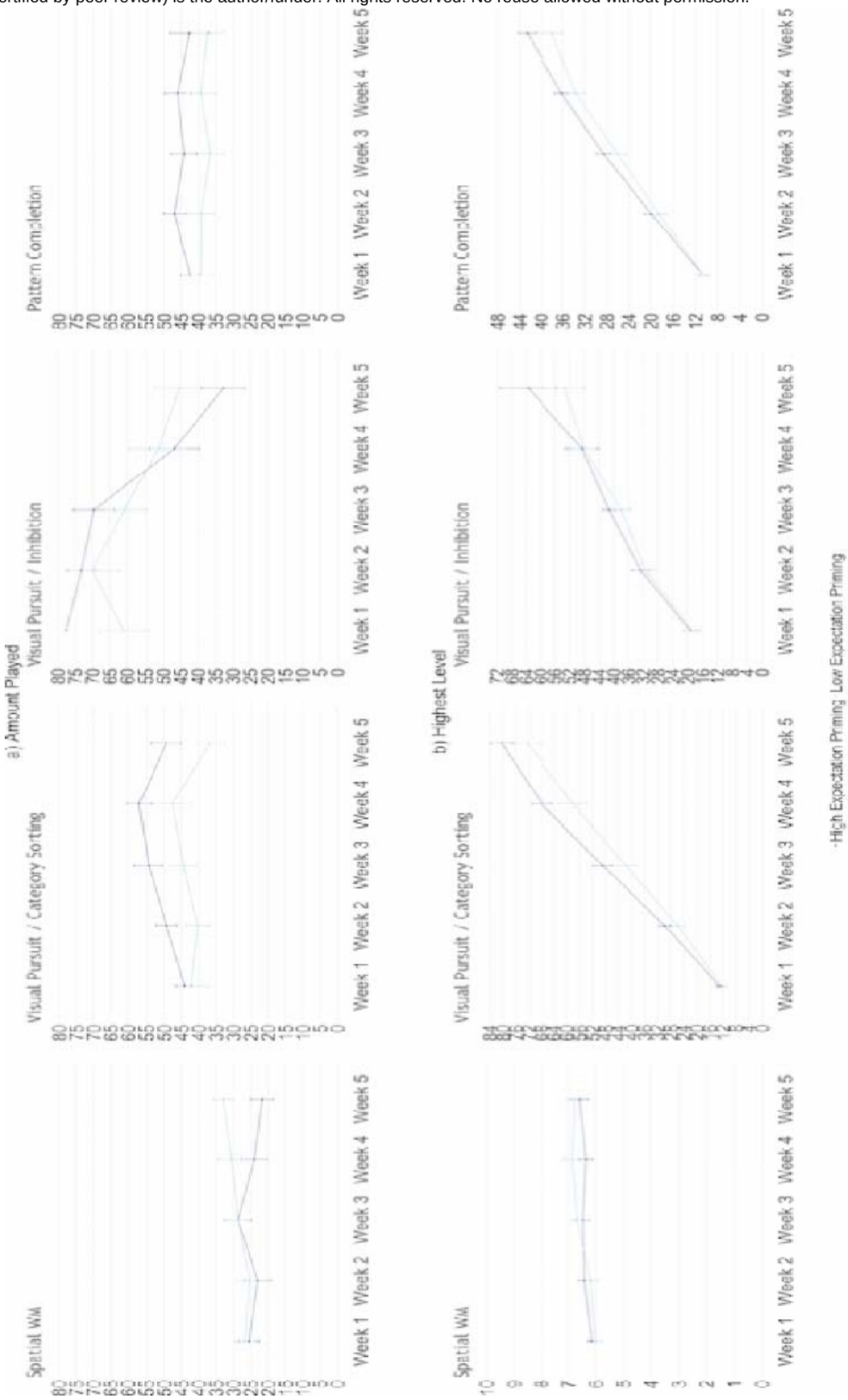

Figure 3. Engagement and performance on the *Activate* games represented by a) the amount of time (min) spent on each game and b) the highest level attained on each game. Error bars represent SEM.

Similarly, for participants in the control program, we used repeated measures MANOVA to

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

examine amount of engagement in each level of Sudoku puzzle and type of n-back exercise over the five weeks (see supplemental files). We found no significant effect of time (Wilk's λ=.22, $F_{(24,12)}$=1.77, p=.152, BF_{10}=.012), of expectation priming (Wilk's λ=.955, $F_{(6,30)}$=.24, p=.961, BF_{10}=.117), or interaction between time and expectation priming (Wilk's λ=.307, $F_{(24,12)}$=1.13, p=.43, BF_{10}=.003).

Performance on Training Tasks

Because we found differences in amount of engagement with the program, we included total amount of time played (i.e., sum of played amount over the five weeks) as a covariate in our repeated measures multivariate analysis of performance on the *Activate* training tasks (Figure 3b). We found a main effect of time on highest level achieved in the *Activate* games (Wilk's λ=.371, $F_{(16,25)}$=2.64, p=.014, η_p^2=.63), but no main effect of expectation priming ($F_{(4,37)}$=1.90, p=.132, BF_{10}=.301) or interaction between time and expectation priming ($F_{(16,25)}$=.91, p=.56, BF_{10}=.01). Follow-up univariate tests revealed the effect of time to be significant for the "spatial WM" ($F_{(2.7,107.1)}$=4.25, p=.009, η_p^2=.096) and "visual pursuit / category sorting" categories ($F_{(1.5,59.9)}$=6.65, p=.005, η_p^2=.14). For the "spatial WM" games, where levels represent list length of items remembered, performance was significantly lower in the second week, compared to the first (p=.041), fourth (p=.029), and fifth weeks (p=.006); for the "visual pursuit / category sorting" games, performance steadily and significantly improved over the five weeks (p<.0001). Note that visual inspection suggests improved performance on the "visual pursuit / category sorting", "visual pursuit / inhibition", and "pattern completion" games over the five weeks. However, based on results form our MANCOVA, these patterns are likely influenced by the amount of time spent on the games.

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

We performed the same analysis to evaluate performance on the control tasks in two separate MANOVAs to account for the two different n-back scores (hit and false alarm rates) analyzed (see supplemental files). For both the number of successfully completed Sudoku puzzles and performance on n-back games, we found no effect of time (Wilk's λ=.828, $F_{(12,30)}$=.519, p=.89, BF_{10}=.023; Wilk's λ=.373, $F_{(24,18)}$=1.26, p=.31, BF_{10}=6.837x10^{-4}, respectively), of expectation priming (Wilk's λ=.938, $F_{(3,39)}$=.866, p=.47, BF_{10}=2.729; Wilk's λ=.977, $F_{(6,36)}$=.144, p=.99, BF_{10}=.309, respectively), or an interaction between the two (Wilk's λ=.652, $F_{(12,30)}$=1.33, p=.25, BF_{10}=.002; Wilk's λ=.493, $F_{(24,18)}$=.773, p=.726, BF_{10}=.002, respectively).

Neuropsychological Transfer Effects

Repeated measures MANOVA examining performance on neuropsychological transfer tests across expectation condition and program revealed a main effect of time (Wilk's λ=.633, $F_{(7,75)}$=6.22, p<.0001, η_p^2=.37) and of expectation priming (Wilk's λ=.761, $F_{(7,75)}$=3.36, p=.004, η_p^2=.24; Figure 4). Notably, we found no significant main effect of program ($BF_{10}\leq$.961) or evidence of interactions between time, program, and expectation priming ($BF_{10}\leq$1.892).

Follow-up analyses indicated a significant post-training improvement in Stroop colour-word score ($F_{(1,81)}$=11.35, p=.001, η_p^2=.12), time on TMT trail B ($F_{(1,81)}$=10.87, p=.001, η_p^2=.12), categories completed on the WCST ($F_{(1,81)}$=9.88, p=.002, η_p^2=.11), letter fluency ($F_{(1,81)}$=7.20, p=.009, η_p^2=.08), and CVLT semantic clustering ($F_{(1,81)}$=4.24, p=.043, η_p^2=.05). Moreover, the effect of expectation was significant for time on TMT Trail B ($F_{(1,81)}$=4.61, p=.035, η_p^2=.05), whereby participants assigned to low expectation priming had faster response times.

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

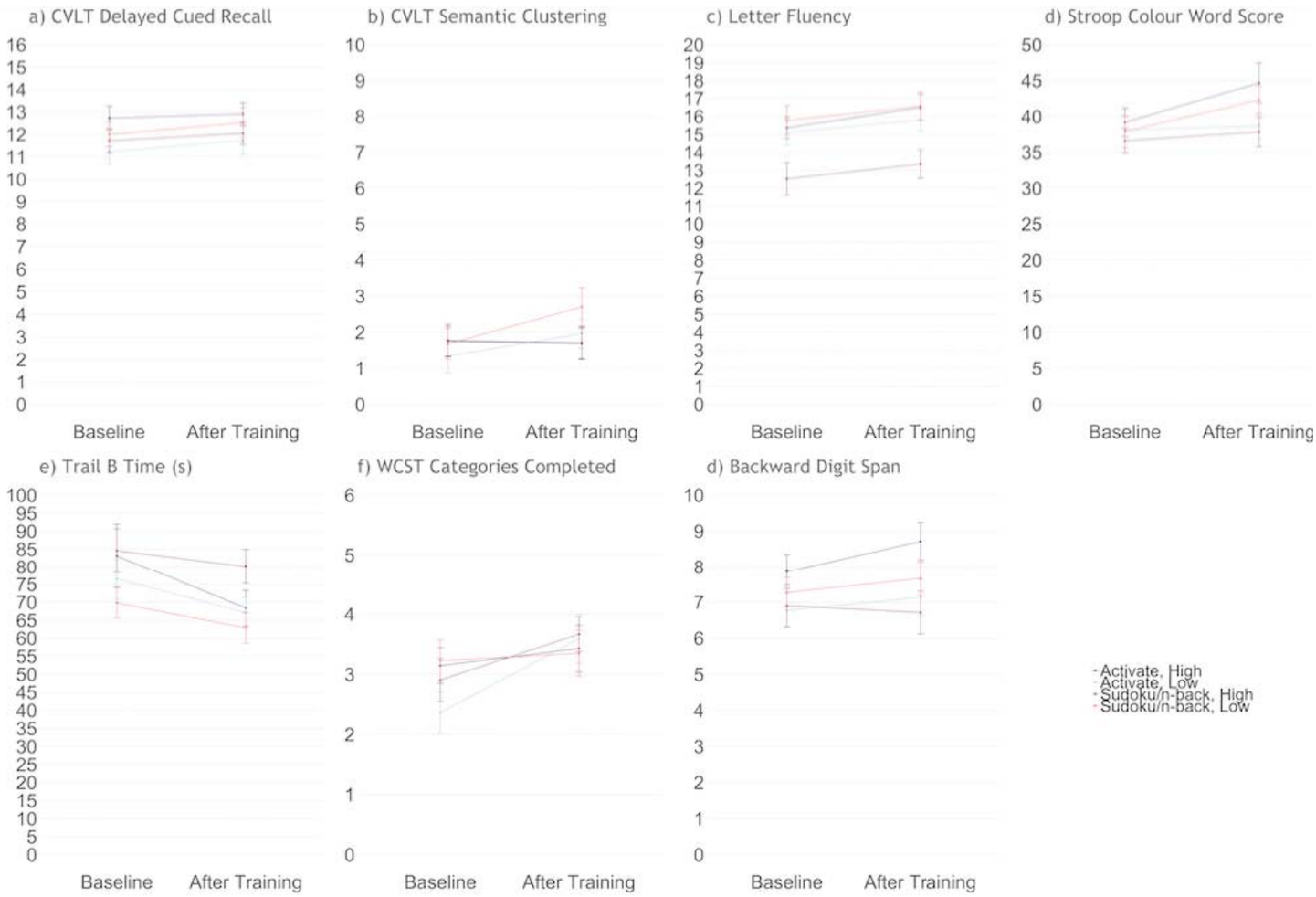

Figure 4. Performance on cognitive tasks analyzed, including: a) long-delay cued recall and b) semantic clustering on the CVLT; c) letter fluency, computed as the average number of words beginning with the letters F, A, and S; d) Stroop colour-word score; e) time (s) on Trail B of the TMT; f) categories completed on the WCST; and g) score on the BDS. Error bars represent SEM.

Discussion

Here we sought to evaluate the effects of expectation priming and a 5-week CCT program on cognitive function and psychosocial wellbeing in healthy older adults. Most participants adhered to their assigned program (88%), believed they were genuinely training their brains (80%) and

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

reported a positive experience overall. All participants improved somewhat on the neuropsychological and psychosocial transfer measures over time, but training on the *Activate* program yielded no greater improvement on the transfer measures than the control training. Similarly, the high expectation priming condition led to no greater improvement (on training or transfer) than the low expectation condition.

Participant Expectations

As in our previous work (Rabipour, Andringa, et al., 2017; Rabipour & Davidson, 2015), here we found that older adults were largely optimistic about CCT outcomes before they began the study. Moreover, ratings at the outset of the program suggested that our expectation priming (i.e., via the EAS) had successfully raised or lowered participant expectations. Following completion of the program, however, participants were more uncertain about the effectiveness of the CCT in improving cognitive outcomes, with ratings closer to the "neutral" score in all experimental groups.

Activate and Control Training

Participants assigned to the *Activate* CCT adhered to the program and improved in three out of four of the games' training domains, including "visual pursuit & category sorting", "visual pursuit & inhibition", and "pattern completion". The relative lack of improvement in the "spatial WM" exercises may reflect the restricted range of these game levels, which represented list length remembered, rather than a combination of correct responses and speed, as with the other categories. The number of days that participants engaged with the program did not differ between groups, although there was considerable variability within groups. Notably, some participants in each of the expectation conditions even continued with the *Activate* program long

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

after they had completed the trial (we did not include data from sessions completed beyond the 5-week trial in our analyses).

Participants who completed the control training (i.e., Sudoku and n-back games) reported being highly engaged with the program, but did not improve on these exercises. Interestingly, participants assigned to the Sudoku and n-back exercises reported that they tended to select a larger number of "hard" Sudoku grids over time, indicating motivation to challenge themselves further. Unfortunately, due to issues with database design, we were unable to analyze performance of the n-back exercises on the basis of level (i.e., 2-vs. 3-vs. 4-back task) here.

We examined amount of logged playtime for both the *Activate* and control games (see Figure 3 and supplemental files) because this was the closest comparison we could make between the two programs. Note, however, that we cannot be certain that the amount of time recorded for the control games consistently reflected engagement in the task, rather than the game remaining open in a web browser after a participant finished, due to reported glitches with the auto-logout. This may explain the relatively large variance for several of the control games. This issue did not appear to exist for the *Activate* games.

Participants assigned to the high expectation priming conditions spent more time on one of the four *Activate* games, compared to their counterparts who received low expectation priming. This suggests that participants in the high expectation conditions may have been more motivated to participate in the program, and may have contributed greater effort throughout the training sessions, at least on some games, compared to those in the low expectation conditions. Nevertheless, lack of significant differences on the training tasks suggests that any possible effects of motivation and effort were minimal, and did not meaningfully contribute to performance.

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(Limited) Transfer Effects Cognitive Function

Although participants provided positive feedback, we did not find any effect of expectation priming or of the CCT program on our transfer measures of cognitive or psychosocial function (see supplemental files). There are several possible reasons.

We recruited a self-selected sample of relatively high-functioning older adults, some of whom who were concerned about their cognitive health and likely to partake in cognitive training on their own. In this regard, our sample is representative of CCT subscribers. Had we recruited more low-functioning older adults or people with psychiatric conditions (e.g., as in Morimoto et al., 2016), we might have found transfer of training. Some promising findings have been reported in people with Mild Cognitive Impairment (MCI), although even in that group the literature is inconsistent (e.g., Butler et al., 2018). One recent high-quality study reported improved memory and strategy use in people with amnesic MCI following eight weekly sessions of group-led cognitive training (Belleville et al., 2018). Nevertheless, there is currently no conclusive evidence that cognitive training can prevent or delay the onset of MCI or other forms of dementia (National Academies of Sciences, 2017).

Although an assessment of strategy use was beyond the scope of this study, it is possible that some participants developed or obtained (e.g., from a friend or family member) more effective strategies on how to complete the games. For example, at post-assessment, a number of participants reported repeating items to be remembered out loud, assigning names to specific repeated stimuli, or placing the cursor in strategic locations in anticipation of a response. This may have limited the benefits of the cognitive training. Similarly, because a large portion of the training was completed from home, we were unable to control for environment during home training sessions, or monitor the potential use of external aids (e.g., other computing software) to complete the Sudoku puzzles or n-back trials.

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Expectation Priming

Regarding the *Activate* training, one possibility for the relative lack of an expectation effect is the subtlety of our priming messages. Although ratings on the EAS appeared to change consistently in the predicted direction in hypothetical and anticipatory scenarios, the messages might be insufficient to alter perceptions when participants have actually undergone an intervention. We elected not to use a more overt or repeated priming strategy to avoid divulging the nature of our expectation manipulation. However, measuring and evaluating the potential influence of expectations on outcomes is difficult, and the most appropriate approach remains debated (Green et al., 2019); it is therefore possible that using other forms of media (e.g., verbal messages or pre-recorded videos) would provide more potent expectation priming. Despite these potential considerations, a recent study using different expectation priming and assessment methods also found little effect of an expectation manipulation (Tsai et al., 2018). One thing to keep in mind in future work is that participants can be quite sophisticated in their ability to recognize or infer the subtleties of experimental design: Many of ours were actively engaged in learning about cognitive aging in their leisure time and participated in research studies as a way to remain informed and involved. Consequently, many stated at the outset or at follow-up that training effects can be nuanced, and that both achieving and measuring transfer can be difficult.

CCT Design

Although we found no significant expectation or training effects on cognition here, we might have under other conditions. For example, it is possible that longer periods of training would lead to measurable effects. The scientific community has yet to agree on an optimal timeframe and intensity of training (i.e., “dose”; Rabipour, Miller, Taler, Messier, & Davidson, 2017).

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Nevertheless, the creators of *Activate* agreed that our training schedule and timeline were reasonable, reaching the minimal 1000 minutes of training recommended to commercial subscribers to detect performance changes. Moreover, increasing either of these parameters (i.e., length or frequency of training) would also increase the risk of participant dropout. Offering booster sessions at a pre-determined time interval (e.g., several months or a year following completion of the program) might help maintain or even enhance training effects (Rebok et al., 2014). Nevertheless, this solution is unlikely to have altered outcomes in the present study given the lack of any immediate benefits. Moreover, although we conducted a 6-month follow-up, we were unable to analyze these data due to low response rate. Such low response rates suggest that few participants would have returned for booster sessions.

Our null findings may also have resulted from using sub-optimal transfer measures. Although we selected our transfer tasks carefully, based on evidence from the literature as well as consensus with the creators of the *Activate* program – researchers and medical professionals at Yale University – it is possible that the behavioural measures were not refined enough to detect small behavioural effects or changes at the neurological level. For this reason, we have begun a follow-up trial of the *Activate* program using identical training and control conditions, but with different behavioural tasks as well as electrophysiological outcome measures. In addition, as with all subjective measures, it is possible that responses on our self-reports of expectations, subjective memory, and wellbeing were unrepresentative (e.g., if participants simply told us what they believed we wanted to hear; i.e., “demand characteristics”). However, reported expectation ratings replicated our previous findings, and the subjective questionnaires we administered are commonly used in studies of cognition, aging, and CCT.

Finally, although we conducted an *a priori* power analysis to determine the minimal required sample size, any effects of this training could be smaller than anticipated and therefore would be

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

more easily discernible with a larger sample size. Increasing the number of participants would enable further exploratory analyses, examining the potential influence of individual factors (e.g., differences in baseline performance, mood, subjective perceptions of performance and function, etc.) in mediating training outcomes. Nevertheless, collapsing across expectation conditions permitted us to evaluate program-specific effects with a larger sub-sample; these effects were non significant.

Conclusions

The present study is the first to directly examine expectations of CCT outcomes, including neuropsychological and psychosocial transfer measures, in the lab setting. By priming participants to have either high or low expectations of training at the outset, we sought to tease apart effects of training from those of expectations of CCT in older adults.

Moreover, we incorporated many of the best practices recommended for high quality intervention research. Crucially, studies rarely account for psychological influences such as expectations of outcomes (Green et al., 2019), a factor we directly evaluated and manipulated experimentally. Examining such factors is particularly important for clinical populations, where placebo effects are known to affect outcomes (Benedetti, Carlino, & Pollo, 2011) and where participants may have more to lose compared to healthy populations. Interestingly, evidence suggests that the placebo effect is weaker, and potentially non-existent, in people with certain forms of dementia (e.g., Alzheimer's; Benedetti et al., 2006). Expectancy-related placebo effects may therefore represent less of a confound in cognitive training trials in this population.

Our findings replicate previous reports that older adults have high expectations of CCT outcomes (Rabipour & Davidson, 2015). Feedback from our participants further supports the importance of credibility on subjective experience, including engagement with the program.

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Nevertheless, our 5-week trial of a commercially available CCT program indicated little effect of training program, or of expectations, on cognitive performance or subjective perceptions of cognition and wellbeing. Our findings may reflect the specific parameters of this study, and do not imply that all CCT programs – even other versions of *Activate* – are ineffective. Nevertheless, these results join with others in suggesting caution regarding the effectiveness of such short-term computerized programs on cognitive and behavioural outcomes (e.g., Kable et al., 2017; Wentink et al., 2016; for a review, see Simons et al., 2016).

Given the widespread appeal of CCT as an affordable and accessible form of cognitive enhancement in both healthy and clinical populations, understanding the potential impact of expectations and of program design is imperative to assessing the practical value of such interventions. Thus, although we found few effects of expectations here, future studies of cognitive intervention should account for participant expectations using a validated tool such as the EAS. Optimizing intervention design and harnessing the potential positive effects of expectations can also be useful in the home, clinic, or other settings. Eventually, this understanding may inform the development of more effective approaches towards enhancing cognitive function and optimizing clinical interventions for the treatment or prevention of cognitive impairment.

Author Contributions

S.R. and P.S.R.D. designed the study with help from M.P. S.R., C.M., and J.C. collected the data. S.R. and P.S.R.D. analyzed the data with help from C.M., J.C., M.O.G.G., and A.P. S.R. and P.S.R.D. wrote the manuscript. All authors reviewed and approved the final version of the manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Acknowledgements

We thank Bruce Wexler and Sean O'Leary for granting us free access to and support for the *Activate* program. We also thank Bruce Wexler, Sarah Morimoto, and Morris Bell for providing feedback on our experimental protocol and analysis plan, as well as Veronika Huta for guidance on our statistical analyses. We thank Bruce Wexler for helpful comments on an earlier draft of this manuscript, as well as Charles Collin, Andra Smith, Vanessa Taler, and Louis Behrer for thoughtful feedback on the study design and analysis. We thank Zachary Ball, Lara Geinoz, Victoria Hall, Alix Hill, Katia Giguère Marshal, Awais Rahman, Raphaëlle Robidoux, Zak Sanaye, Jonathan Tran, and Allison Walsh for help with data collection and verification. We also acknowledge the Natural Sciences and Engineering Research Council (NSERC, Discovery Grant), Ontario Graduate Scholarships, and Fonds de Recherche Québec – Santé for their support of this work.

Conflict of interest statement

On behalf of all authors, the corresponding author states that there is no conflict of interest.

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

Anderson, N. D., & Craik, F. I. (2017). 50 Years of Cognitive Aging Theory. *J Gerontol B Psychol Sci Soc Sci, 72*(1), 1-6. doi: 10.1093/geronb/gbw108

Belleville, S., Hudon, C., Bier, N., Brodeur, C., Gilbert, B., Grenier, S., . . . Gauthier, S. (2018). MEMO plus : Efficacy, Durability and Effect of Cognitive Training and Psychosocial Intervention in Individuals with Mild Cognitive Impairment. *Journal of the American Geriatrics Society, 66*(4), 655-663. doi: 10.1111/jgs.15192

Benedetti, F., Arduino, C., Costa, S., Vighetti, S., Tarenzi, L., Rainero, I., & Asteggiano, G. (2006). Loss of expectation-related mechanisms in Alzheimer's disease makes analgesic therapies less effective. *Pain, 121*(1-2), 133-144. doi: 10.1016/j.pain.2005.12.016

Benedetti, F., Carlino, E., & Pollo, A. (2011). How Placebos Change the Patient's Brain. *Neuropsychopharmacology, 36*(1), 339-354. doi: Doi 10.1038/Npp.2010.81

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. doi: 10.1177/1745691613491271

Boot, W. R., Souders, D. J., Charness, N., Blocker, K., Roque, N. A., & Vitale, T. (2016). The Gamification of Cognitive Training: Older Adults' Perceptions of and Attitudes Toward Digital Game-Based Interventions. In J. Zhou & G. Salvendy (Eds.), *Human Aspects of IT for the Aged Population. Design for Aging. ITAP* (Vol. 9754, pp. 290-300): Springer, Cham.

Butler, M., McCreedy, E., Nelson, V. A., Desai, P., Ratner, E., Fink, H. A., . . . Kane, R. L. (2018). Does Cognitive Training Prevent Cognitive Decline?: A Systematic Review. *Ann Intern Med, 168*(1), 63-68. doi: 10.7326/M17-1531

Campbell, N. L., Maidment, I., Fox, C., Khan, B., & Boustani, M. (2013). The 2012 update to the anticholinergic cognitive burden scale. *Journal of the American Geriatrics Society, 61*(S1), S142–S143. doi: 10.1111/jgs.2013.61.issue-s1

Collie, A., Maruff, P., Darby, D. G., & McStephen, M. (2003). The effects of practice on the cognitive test performance of neurologically normal individuals assessed at brief test-retest intervals. *J Int Neuropsychol Soc, 9*(3), 419-428. doi: 10.1017/S1355617703930074

Crowe, S. F. (2000). Does the letter number sequencing task measure anything more than digit span? *Assessment, 7*(2), 113-117. doi: Doi 10.1177/107319110000700202

Davidson, P. S. R., & Winocur, G. (2017). Aging and Cognition. *Reference Module in Neuroscience and Biobehavioral Psychology*, 1-7. doi: 10.1016/B978-0-12-809324-5.00252-2

Drag, L. L., & Bieliauskas, L. A. (2010). Contemporary review 2009: cognitive aging. *J Geriatr Psychiatry Neurol, 23*(2), 75-93. doi: 10.1177/0891988709358590

Foroughi, C. K., Monfort, S. S., Paczynski, M., McKnight, P. E., & Greenwood, P. M. (2016). Placebo effects in cognitive training. *Proc Natl Acad Sci U S A, 113*(27), 7470-7474. doi: 10.1073/pnas.1601243113

Glisky, E. L., Polster, M. R., & Routhieaux, B. C. (1995). Double Dissociation Between Item and Source Memory. *Neuropsychology, 9*(2), 229-235. doi: 10.1037/0894-4105.9.2.229

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*, 1-28. doi: https://doi.org/10.1007/s41465-018-0115-y

Heinzel, S., Schulte, S., Onken, J., Duong, Q. L., Riemer, T. G., Heinz, A., . . . Rapp, M. A. (2014). Working memory training improvements and gains in non-trained cognitive tasks

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

in young and older adults. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn, 21*(2), 146-173. doi: 10.1080/13825585.2013.790338

Jaeggi, S. M., Buschkuehl, M., Shah, P., & Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Memory & Cognition, 42*(3), 464-480. doi: 10.3758/s13421-013-0364-z

Kable, J. W., Caulfield, M. K., Falcone, M., McConnell, M., Bernardo, L., Parthasarathi, T., . . . Lerman, C. (2017). No Effect of Commercial Cognitive Training on Brain Activity, Choice Behavior, or Cognitive Performance. *J Neurosci, 37*(31), 7390-7402. doi: 10.1523/JNEUROSCI.2832-16.2017

Kavanaugh, B. C., Tuncer, O. F., & Wexler, B. E. (2018). Measuring and Improving Executive Functioning in the Classroom. *Journal of Cognitive Enhancement*, 1-10. doi: 10.1007/s41465-018-0095-y

Kortte, K. B., Horner, M. D., & Windham, W. K. (2002). The trail making test, part B: cognitive flexibility or ability to maintain set? *Appl Neuropsychol, 9*(2), 106-109. doi: 10.1207/S15324826AN0902_5

Lamar, M., Resnick, S. M., & Zonderman, A. B. (2003). Longitudinal changes in verbal memory in older adults - Distinguishing the effects of age from repeat testing. *Neurology, 60*(1), 82-86.

Lampit, A., Hallock, H., & Valenzuela, M. (2014). Computerized cognitive training in cognitively healthy older adults: a systematic review and meta-analysis of effect modifiers. *PLoS Med, 11*(11), e1001756. doi: 10.1371/journal.pmed.1001756

Lansing, A. E., Ivnik, R. J., Cullum, C. M., & Randolph, C. (1999). An empirically derived short form of the Boston naming test. *Arch Clin Neuropsychol, 14*(6), 481-487.

Lovden, M., Backman, L., Lindenberger, U., Schaefer, S., & Schmiedek, F. (2010). A theoretical framework for the study of adult cognitive plasticity. *Psychol Bull, 136*(4), 659-676. doi: 10.1037/a0020080

McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The Relationship Between Working Memory Capacity and Executive Functioning: Evidence for a Common Executive Attention Construct. *Neuropsychology, 24*(2), 222-243. doi: 10.1037/a0017619

Morimoto, S. S., Gunning, F. M., Wexler, B. E., Hu, W., Ilieva, I., Liu, J., . . . Alexopoulos, G. S. (2016). Executive Dysfunction Predicts Treatment Response to Neuroplasticity-Based Computerized Cognitive Remediation (nCCR-GD) in Elderly Patients with Major Depression. *Am J Geriatr Psychiatry, 24*(10), 816-820. doi: 10.1016/j.jagp.2016.06.010

Morimoto, S. S., Wexler, B. E., Liu, J. C., Hu, W. L., Seirup, J., & Alexopoulos, G. S. (2014). Neuroplasticity-based computerized cognitive remediation for treatment-resistant geriatric depression. *Nature Communications, 5*. doi: 10.1038/Ncomms5579

Morrison, A. B., & Chein, J. M. (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin & Review, 18*(1), 46-60. doi: Doi 10.3758/S13423-010-0034-0

Motter, J. N., Devanand, D. P., Doraiswamy, P. M., & Sneed, J. R. (2016). Clinical Trials to Gain FDA Approval for Computerized Cognitive Training: What Is the Ideal Control Condition? *Frontiers in Aging Neuroscience, 8*. doi: 10.3389/fnagi.2016.00249

Nasreddine, Z. S., Chertkow, H., Phillips, N., Whitehead, V., Collin, I., & Cummings, J. L. (2004). The Montreal Cognitive Assessment (MoCA): a brief cognitive screening tool for detection of Mild Cognitive Impairment. *Neurology, 62*(7), A132-A132.

National Academies of Sciences, E., and Medicine. (2017). *Preventing Cognitive Decline and Dementia: A Way Forward*. Washington (DC).

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Rabipour, S., Andringa, R., Boot, W. R., & Davidson, P. S. R. (2017). What Do People Expect of Cognitive Enhancement? *Journal of Cognitive Enhancement*, 1-8. doi: 10.1007/s41465-017-0050-3

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. doi: 10.1016/j.bbr.2015.01.002

Rabipour, S., Davidson, P. S. R., & Kristjansson, E. (2018). Measuring Expectations of Cognitive Enhancement: Item Response Analysis of the Expectation Assessment Scale. *Journal of Cognitive Enhancement*. doi: 10.1007/s41465-018-0073-4

Rabipour, S., Miller, D., Taler, V., Messier, C., & Davidson, P. S. R. (2017). Physical and Cognitive Exercise in Aging. In L. Rigby (Ed.), *Handbook of Gerontology Research Methods: Understanding Successful Aging*: Routledge.

Rebok, G. W., Ball, K., Guey, L. T., Jones, R. N., Kim, H. Y., King, J. W., . . . Grp, A. S. (2014). Ten-Year Effects of the Advanced Cognitive Training for Independent and Vital Elderly Cognitive Training Trial on Cognition and Everyday Functioning in Older Adults. *Journal of the American Geriatrics Society, 62*(1), 16-24. doi: Doi 10.1111/Jgs.12607

Redick, T. S. (2015). Working memory training and interpreting interactions in intelligence interventions. *Intelligence, 50*, 14-20. doi: 10.1016/j.intell.2015.01.014

Redick, T. S., & Lindsey, D. R. B. (2013). Complex span and n-back measures of working memory: A meta-analysis. *Psychonomic Bulletin & Review, 20*(6), 1102-1113. doi: 10.3758/s13423-013-0453-9

Reuter-Lorenz, P. A., Festini, S. B., & Jantz, T. K. (2016). Executive Functions and Neurocognitive Aging *Handbook of the Psychology of Aging* (8th ed., pp. 245-262): Academic Press.

Rhodes, M. G. (2004). Age-related differences in performance on the Wisconsin card sorting test: a meta-analytic review. *Psychol Aging, 19*(3), 482-494. doi: 10.1037/0882-7974.19.3.482

Rohsenow, D. J., & Marlatt, G. A. (1981). The Balanced Placebo Design - Methodological Considerations. *Addictive Behaviors, 6*(2), 107-122. doi: Doi 10.1016/0306-4603(81)90003-4

Shah, T. M., Weinborn, M., Verdile, G., Sohrabi, H. R., & Martins, R. N. (2017). Enhancing Cognitive Functioning in Healthly Older Adults: a Systematic Review of the Clinical Significance of Commercially Available Computerized Cognitive Training in Preventing Cognitive Decline. *Neuropsychol Rev, 27*(1), 62-80. doi: 10.1007/s11065-016-9338-9

Shipstead, Z., Redick, T. S., & Engle, R. W. (2010). Does Working Memory Training Generalize? *Psychologica Belgica, 50*(3-4), 245-276.

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is Working Memory Training Effective? *Psychological Bulletin, 138*(4), 628-654. doi: Doi 10.1037/A0027473

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. L. (2016). Do "Brain-Training" Programs Work? *Psychological Science in the Public Interest, 17*(3), 103-186. doi: 10.1177/1529100616661983

Souders, D. J., Boot, W. R., Blocker, K., Vitale, T., Roque, N. A., & Charness, N. (2017). Evidence for Narrow Transfer after Short-Term Cognitive Training in Older Adults. *Frontiers in Aging Neuroscience, 9*. doi: 10.3389/fnagi.2017.00041

Spieler, D. H., Balota, D. A., & Faust, M. E. (1996). Stroop performance in healthy younger and older adults and in individuals with dementia of the Alzheimer's type. *Journal of Experimental Psychology-Human Perception and Performance, 22*(2), 461-479. doi: Doi 10.1037/0096-1523.22.2.461

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Stricker, J. L., Brown, G. G., Wixted, J., Baldo, J. V., & Delis, D. C. (2002). New semantic and serial clustering indices for the California Verbal Learning Test-Second Edition: background, rationale, and formulae. *J Int Neuropsychol Soc, 8*(3), 425-435.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Toepper, M. (2017). Dissociating Normal Aging from Alzheimer's Disease: A View from Cognitive Neuroscience. *J Alzheimers Dis, 57*(2), 331-352. doi: 10.3233/JAD-161099

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*. doi: doi.org/10.1016/j.jarmac.2018.06.001

Wadley, V. G., Benz, R. L., Ball, K. K., Roenker, D. L., Edwards, J. A., & Vance, D. E. (2006). Development and evaluation of home-based speed-of-processing training for older adults. *Archives of Physical Medicine and Rehabilitation, 87*(6), 757-763. doi: Doi 10.1016/J.Apmr.2006.02.027

Wentink, M. M., Berger, M. A. M., de Kloet, A. J., Meesters, J., Band, G. P. H., Wolterbeek, R., . . . Vlieland, T. P. M. V. (2016). The effects of an 8-week computer-based brain training programme on cognitive functioning, QoL and self-efficacy after stroke. *Neuropsychological Rehabilitation, 26*(5-6), 847-865. doi: 10.1080/09602011.2016.1162175

Zanto, T. P., Chadick, J. Z., & Gazzaley, A. (2014). Anticipatory alpha phase influences visual working memory performance. *Neuroimage, 85*, 794-802. doi: Doi 10.1016/J.Neuroimage.2013.07.048

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplemental Files:

Few Effects of a 5-week Computerized Cognitive Training Program in Healthy Older Adults

Sheida Rabipour[1], Cassandra Morrison[1,2], Jessica Crompton[1], Marcelo Petrucelli[3], Murillo de Oliveira Gonçalves Germano[4], Anita Popescu[3], and Patrick S. R. Davidson[1,2]

Supplemental Methods

We included subjective measures in addition to our objective tests to permit a more comprehensive analysis of potential training-related improvements on untrained tasks and general functioning.

Psychosocial Indices

Multifactorial Memory Questionnaire (MMQ). The MMQ (Troyer and Rich, 2018) is a metamemory questionnaire that includes scales of contentment or satisfaction (i.e., feelings regarding memory performance), ability (i.e., self-appraisal of memory capabilities), and strategy (i.e., reported frequency of memory strategy use). Self-report of memory may be an important predictor of cognitive decline (Reid and Maclullich, 2006). Nevertheless, both healthy older adults and those with memory disorders can report inaccurate beliefs about memory and aging – which may reflect negative cultural stereotypes – and poor evaluation of their own memory (Troyer and Rich, 2002). Such inaccuracies may lower expectations, reduce effort on everyday tasks involving memory, and lead to impaired memory performance.

Pittsburgh Sleep Quality Index (PSQI). The PSQI is a brief scale measuring sleep quality. Studies suggest that sleep quality may deteriorate in healthy older adults, independent of medication use or the presence of other physical or psychological conditions (Buysse et al., 1991; Nebes et al., 2009). Poor quality of sleep has been shown to associate with greater risk of cognitive decline and dementia (Spira et al., 2014).

Subjective Age Questionnaire. We included a measure of subjective age, adapted from (Hughes et al., 2013), which asks participants to rate their subjective age in years on four categories: i) general age; ii) age with respect to wellbeing; iii) physical age; and iv) cognitive

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

age. Studies suggest that older adults feel older after immediately performing memory tests (Hughes et al., 2013), and that perceived age may correlate with cognitive function (Stephan et al., 2016). We sought to determine whether this pattern would persist in our sample or, conversely, whether perceived age would decrease after completing the cognitive intervention. We used a computerized version of the scale, asking participants to indicate their perceived age in each category by dragging the cursor to the appropriate location on a line denoted only by a "0" and "120" on either end. The relative location of the cursor determined subjective age.

Patient Health Questionnaire (PHQ-9). The PHQ-9 is a brief questionnaire that assesses the presence and severity of depression. PHQ-9 reports correlate with mental and general health, as well as social and physical functioning (Kroenke et al., 2001). We included the PHQ-9 in our final 71 participants based on advice from our colleagues, as a more sensitive substitute for the Center for Epidemiological Studies-Depression (CES-D) scale (Lewinsohn et al., 1997).

Social Engagement Questionnaire (SEQ). The SEQ measures indices of social engagement, including social activity, social network size, and social support (Krueger et al., 2009). We included this measure in our final 53 participants based on feedback from our colleagues, and slightly adapted the form under the guidance of Kristin Krueger, who initially developed the scale.

Short Form Health Survey (SF-36). The SF-36 measure of health-related QOL is reliable in older adults (Hayes et al., 1995). The questionnaire categorizes the items into nine scales, including physical functioning, physical health problems, emotional problems, vitality, mental health, social functioning, bodily pain, general health, and health transition.

Feedback on Perceived Experience

At the post-training assessment, we asked participants to complete a satisfaction questionnaire for feedback regarding their perceived experience during the intervention. On a scale from 1-7 (1 = "very strongly disagree," 2 = "strongly disagree," 3 = "disagree," 4 = neutral/"neither agree nor disagree," 5 = "agree," 6 = "strongly agree," 7 = "very strongly agree"), participants rated the degree to which they found the experience: i) enjoyable; ii) challenging; iii) frustrating; iv) engaging; v) boring; vi) motivating; and vii) satisfying.

Supplemental Results

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Psychosocial Transfer Effects

Repeated measures MANOVA examining subjective reports of psychosocial data across program and expectation condition revealed neither a main effect of time, program, or expectation priming ($F_{(5,77)} \leq 2.11$, $p \geq .073$, BF10≤.208), nor any interactions between them ($F_{(5,77)} \leq 1.81$, $p \geq .122$, BF10≤.364; Figure 6).

Subjective Experience

The majority of participants (70/87=80%) believed their assigned program was genuinely training their brain. Interestingly, the proportion of participants with this belief was larger, albeit not significantly, in the active control condition (38/43=88%) compared to the Activate intervention (32/44=73%; $X^2=4.7$, p=.066).

Participants provided largely positive feedback and reported that they would recommend the program (Figure 7a), regardless of experimental condition (77/87=89%; $X^2=2.45$, ns). MANOVA of participant feedback across expectation priming and CCT program revealed a marginal effect of program (Wilk's $\lambda=.838$, $F_{(7,77)}=2.12$, p=.051, $\eta_p^2=.162$), but not of expectation priming ($F_{(7,77)}=.939$, p=.482) or an interaction ($F_{(7,77)}=.714$, p=.66), on subjective experience. After correcting for multiple comparisons, participants in the active control group (i.e., training on Sudoku and n-back tasks) reported significantly higher enjoyment ($F_{(1,83)}=7.48$, p=.008, $\eta_p^2=.083$) and lower boredom ($F_{(1,83)}=9.42$, p=.003, $\eta_p^2=.102$) of the program.

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

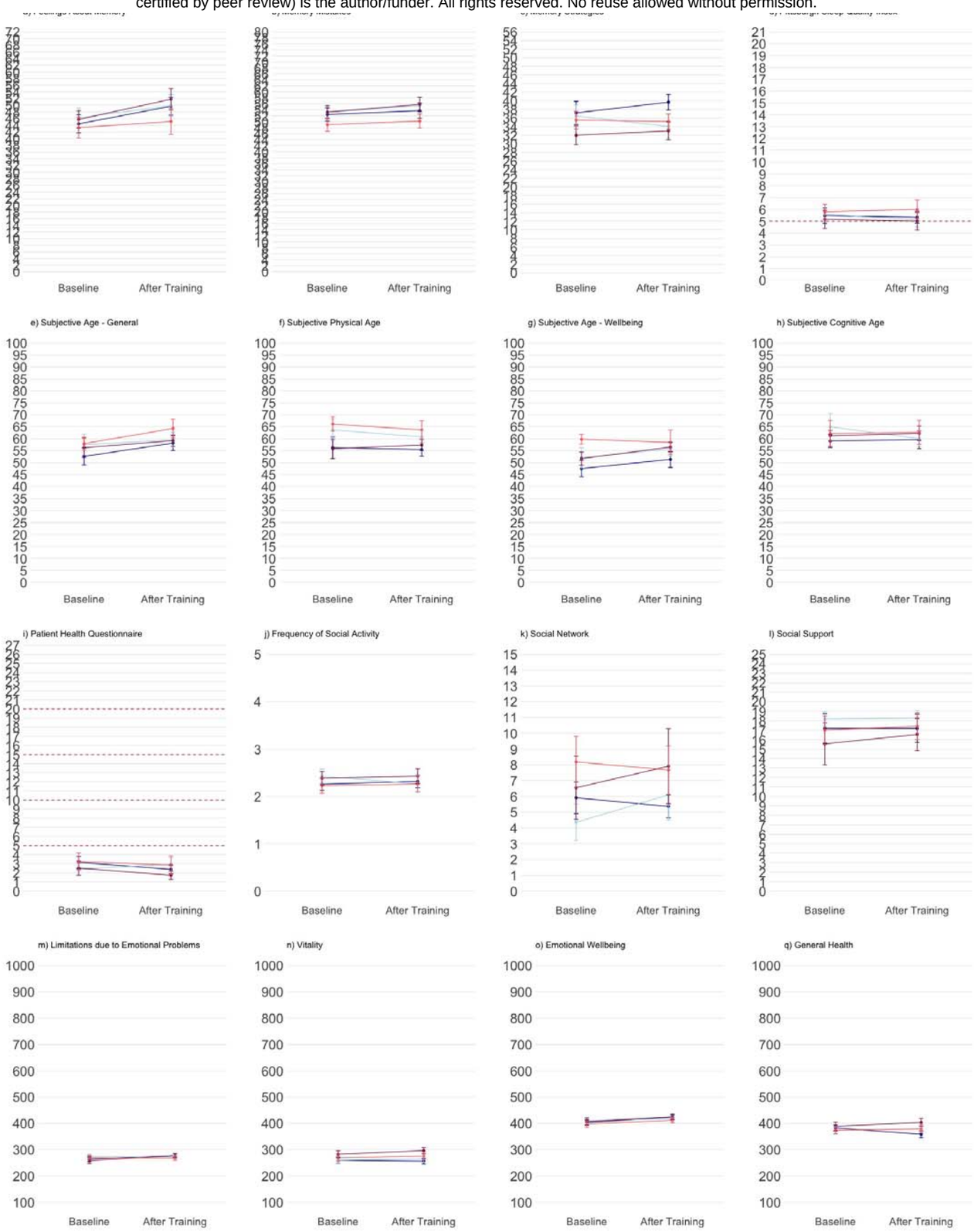

Figure 6. Subjective reports of (a-c) memory, (d) sleep quality, (e-h) perceived age, (i) mood, (j-l) social engagement, and (m-p) quality of life. Solid blue lines represent Activate groups; solid red lines represent control groups. Darker lines represent high expectation priming groups; lighter lines represent low expectation priming groups. Dashed red lines represent clinical threshold(s), where applicable. Error bars represent SEM.

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Perhaps more interestingly, MANOVA of feedback ratings revealed that believing the program genuinely trained the brain significantly influenced subjective experience (Wilk's λ=.751, $F_{(7,79)}$=3.75, p=.001, η_p^2=.25; Figure 7b). Specifically, after correcting for multiple comparisons, people who perceived their assigned program as having genuinely trained their brain rated higher enjoyment ($F_{(1,85)}$=13.47, p≤.0001, η_p^2=.137) and engagement ($F_{(1,85)}$=12.42, p=.001, η_p^2=.127) with their assigned program, and lower boredom ($F_{(1,85)}$=18.70, p<.0001, η_p^2=.180). Note that, due to an unequal distribution of "believers" vs. "non-believers" and, in particular, low number of "non-believers", we collapsed "belief" (i.e., perception of training) across experimental conditions to perform these analyses.

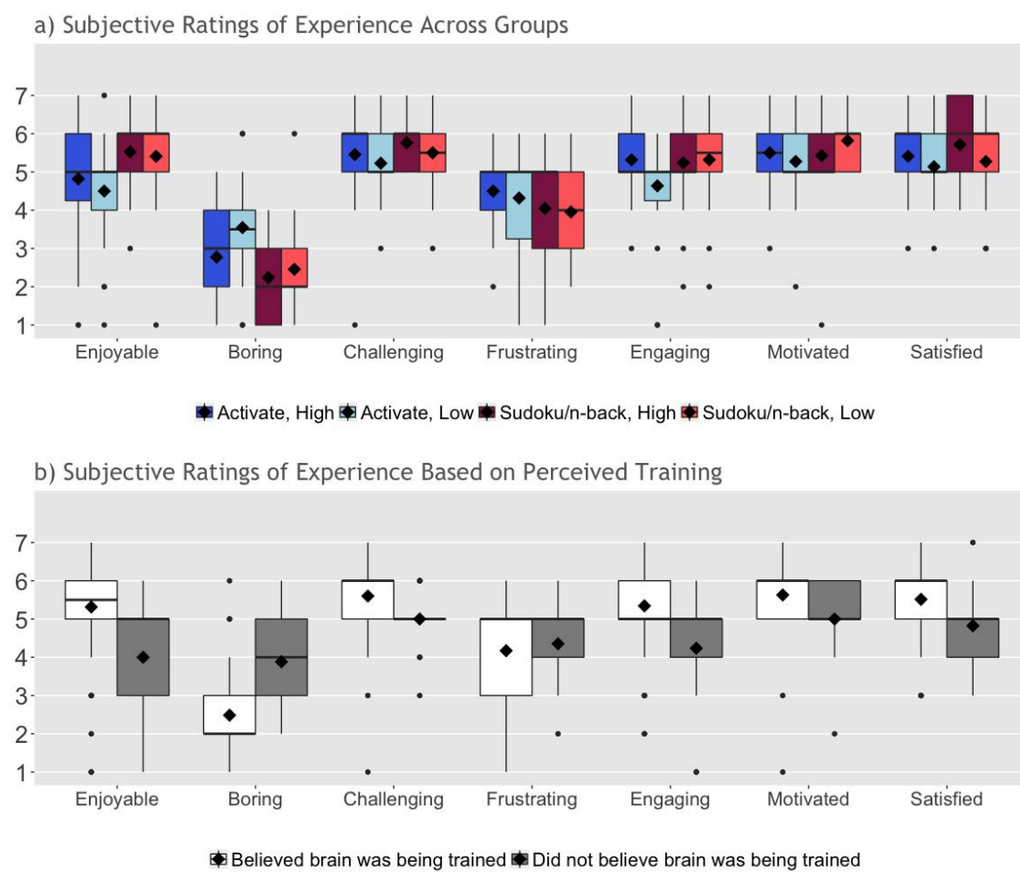

Figure 7. Subjective feedback on experience based on a) experimental group and b) belief that the assigned program genuinely trained the brain. Feedback ratings were on a scale of "1" ("completely disagree") to "7" ("completely agree") for each subscale. Solid lines represent group medians; diamonds represent group means.

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Long Term Outcomes

About half of participants completed at least part of the 6-month follow-up assessment (n=53 responses on the psychosocial questionnaires; n=38 responses on the EAS, final feedback questionnaire, and informal follow-up interview). Due to low power, we elected not to analyze these outcomes.

Supplemental Discussion

Psychosocial Function

We did not detect any changes in self-reported measures of cognitive performance, psychosocial wellbeing, or health-related QOL. This finding may represent the – perhaps unrepresentative – nature of our participant sample, which may have comprised particularly high functioning older adults. Notably, the older adults we recruited were highly educated, independent, and motivated members of the Ottawa community. Many participants demonstrated considerable insight into their own behaviour as well as our study design and evaluation of the program. MoCA scores suggest that our participants had high global cognition, with means in all experimental conditions above the clinical cutoff of "26/30". Moreover, participants tended to have high subjective perceptions of cognition, wellbeing, and QOL at baseline. Variability in these measures, however, may moderate the effects of cognitive training (e.g., people with low ratings of affect or QOL may benefit from CCT to a greater extent; Morimoto et al., 2016). Although we did not find large changes in these outcomes, future studies with greater sample sizes may consider analyzing these variables as mediators or predictors rather than direct outcomes, as we did here.

Subjective Perceptions

Participants enjoyed both the *Activate* and control (Sudoku/n-back) program. Interestingly, participants seemed to enjoy the Sudoku and n-back exercises more than the *Activate* games, and more participants assigned to those exercises believed the program was genuinely training the brain compared to participants who completed *Activate* training. Moreover, we found greater enjoyment and engagement, as well as lower boredom in participants who believed the program had genuinely trained their brain. This finding suggests that, similar to other contexts (e.g., in

bioRxiv preprint doi: https://doi.org/10.1101/570143; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

medicine; Gollub and Kong, 2011; Wampold and Imel, 2007), belief in the capacity of a CCT program to achieve its intended goal (i.e., train the brain) can significantly impact behavioural and other outcomes. Although we did not have a large enough sample of "believers" vs. "non-believers" across our experimental conditions to probe this effect further, future studies may wish to examine how such retrospective perceptions of a program's effectiveness may interact with other outcomes.