Targeted training: Converging evidence against the transferable benefits of online brain training on cognitive function

Bobby Stojanoski*, Kathleen M. Lyons, Alexandra A.A. Pearce, Adrian M. Owen

The Brain and Mind Institute, Western University, London, ON, Canada N6A 3K7

A B S T R A C T

There is strong incentive to improve our cognitive abilities, and brain training has emerged as a promising approach for achieving this goal. While the idea that extensive 'training' on computerized tasks will improve general cognitive functioning is appealing, the evidence to support this remains contentious. This is, in part, because of poor criteria for selecting training tasks and outcome measures resulting in inconsistent definitions of what constitutes transferable improvement to cognition. The current study used a targeted training approach to investigate whether training on two different, but related, working memory tasks (across two experiments, with 72 participants) produced transferable benefits to similar (quantified based on cognitive and neural profiles) untrained test tasks. Despite significant improvement on both training tasks, participants did not improve on either test task. In fact, performance on the test tasks after training were nearly identical to a passive control group. These results indicate that, despite maximizing the likelihood of producing transferable benefits, brain training does not generalize, even to very similar tasks. Our study calls into question the benefit of cognitive training beyond practice effects, and provides a new framework for future investigations into the efficacy of brain training.

1. Introduction

The prospect of enhancing our cognitive abilities is alluring, and there is good incentive to want to do so. Performance on measures of different aspects of cognition, such as processing speed, reasoning, and general intelligence have not only been linked to academic and professional success, but also to happiness, and even life expectancy (Calvin et al., 2011). While cognitive abilities tend to remain relatively stable throughout the lifespan, they are not immune to fluctuations; disease (Marinus et al., 2003; Muller et al., 2007), head injuries (Bleiberg et al., 2004), even at a young age (Talavage et al., 2014), and aging can all result in substantial impairments to cognition. However, the trajectory for cognitive change is not always downward; for example, learning through education or practice is clearly one way in which cognition can be enhanced, and have long lasting effects (Ritchie et al., 2013). However, the cognitive benefits associated with education often progress slowly, require significant investment, and unfold over a long period of time. Recently, brain (or cognitive) training has emerged as a potential new approach for improving cognition – one that is easily accessible and can occur on a much shorter time scale. Moreover, the purported benefits of brain training are not limited to improving cognition, but may include therapeutic benefits that slow, or even reverse,

cognitive decline across the lifespan (Anguera et al., 2013; Westerberg et al., 2007).

Brain training rests on the assumption that regular and prolonged "training" on computerized tasks (often marketed as "brain games") will result in improvements, not only on the trained task, but also on untrained (and even unrelated) tasks, across different cognitive domains. The focus of many brain training programs is on short-term (working) memory - the ability to hold and manipulate information (Baddeley, 1992) – because short-term (working) memory is considered to be the critical cognitive domain underlying generalizable gains in cognition. This notion rests on two key assumptions: 1) that short-term (working) memory can be improved (Klingberg, 2010), and 2) that short-term (working) memory is closely related to other higher-order cognitive abilities, such as, attention (Klingberg et al., 2005), reasoning, problem solving, executive processes (Kane et al., 2004; McCabe et al., 2010; Süß et al., 2002), multitasking (Redick et al., 2016), and even general intelligence (Engle et al., 1999; Kane and Engle, 2002). The logic is intuitive and appealing; brain training programs that increase short-term (working) memory capacity will lead to performance gains across a variety of other cognitive abilities associated with short-term (working) memory (Klingberg, 2010) including general intelligence (see Redick et al., 2013 for evidence why this logic is limited). The idea

* Corresponding author.
  E-mail addresses: bstojan@uwo.ca (B. Stojanoski), klyons8@uwo.ca (K.M. Lyons), apearc4@uwo.ca (A.A.A. Pearce), uwocerc@uwo.ca (A.M. Owen).

has also gained some empirical support in recent years.

For example, several studies have claimed to show that training short-term (working) memory produces generalizable improvements in cognition across various untrained tasks, each measuring different aspects of cognition (Jaeggi et al., 2008). The observed benefits range from improvements on variants of the same task (Li et al., 2008), to improvements on similar tasks that rely on overlapping cognitive mechanisms (*near transfer*; Chein and Morrison, 2010; Dahlin et al., 2008; Tulbure and Siberescu, 2013), to performance gains on unrelated cognitive tasks and domains (*far transfer*;(Au et al., 2014; Caeyenberghs et al., 2016; Chein and Morrison, 2010; Jaeggi et al., 2008; Morrison and Chein, 2010). In fact, it has been suggested that cognitive training can have far reaching consequences, including improvements at work and school in activities such as reading (Dahlin, 2011; Swanson and Jerman, 2006) and math proficiency (Bergman-Nutley and Klingberg, 2014). It has also been claimed that brain training can delay aging-related cognitive decline and reduce the effects of cognitive disease (Basak et al., 2008).

However, the efficacy of brain training has recently been called into question (Simons et al., 2016). For example, some attempts to replicate earlier findings showing brain training-related benefits have failed to produce similar effects (Redick et al., 2013; Thompson et al., 2013). Moreover, one large scale meta-analysis that included studies using multiple forms of short-term (working) memory training found no convincing evidence of transfer of benefits (near and far) to untrained tasks (Melby-Lervåg et al., 2016). In fact, difficulties in finding such transfer effects are not limited to short-term (working) memory based training protocols, but extend to training involving inhibitory control (Enge et al., 2014), video game playing (Lee et al., 2012), and decision making (Kable et al., 2017). In one large-scale study involving more than 11,000 participants, Owen et al. (2010) had participants train for six weeks on a variety of tasks based on commercially available brain training games. They found that, while performance improved on every trained task, there were no gains in performance on untrained tests of reasoning, verbal abilities or short-term memory.

A number of reasons have been proposed to account for these failures to reproduce the results of earlier brain-training studies, including participant's expectations (Foroughi et al., 2016), neuroanatomical variability (Simon et al., 2016), and methodological factors, such as, different analysis approaches (Redick et al., 2013). However, a more fundamental issue likely underlying the variability across studies relates to inconsistent and often vague definitions of what constitutes 'transfer'. The terms 'near' and 'far' transfer are often used to refer to improvements in closely related and unrelated cognitive tasks, respectively, yet how 'related' one task really is to another is often poorly understood. In fact, the degree to which the training tasks differ from the test tasks (and the test tasks from each other) is rarely quantified, and tasks are often selected based on their inferred cognitive properties, rather than some empirical measure of similarity. Without a consistent definition of transfer, and quantifiable measures of similarity between tasks, it is very difficult to make comparisons across studies, and assess the reliability of any observed training related benefits.

To provide a more constrained framework for brain training, the current study focused on two fundamental, but related issues: the nature of the training protocol, and the selection of the tests themselves. Two experiments were conducted that employed a targeted training protocol; in each experiment, participants trained extensively on only one task (unique to each experiment) measuring a single cognitive domain – short-term (working) memory. In addition, quantifiable measures of similarity were used to guide the selection of test tasks. The training and test tasks were taken from the Cambridge Brain Sciences (CBS) battery, an online suite of 12 cognitive assessment tools. One short-term (working) memory task that involved memory for spatial locations was selected for training in experiment 1. Two other tasks were selected to assess the benefits of transfer, one that also involved spatial working memory and one that was procedurally similar, yet

involved verbal working memory. These selections were made based on quantifiable measures of similarity using a factor analysis of task performance and underlying neural activity (Hampshire et al., 2012).

To ensure the results were generalizable, in experiment 2, a short-term (working) memory task that has been widely used in brain training studies was selected for training. The dual n-back task shares many of the same cognitive and neural properties as the task that was selected for training in experiment 1 (Owen et al., 2005) and has successfully produced both near and far transfer in previous studies (Au et al., 2016; Jaeggi et al., 2008). In experiment 2, the same two spatial and verbal working memory tasks (that were used in experiment 1) were used to assess the effects of training.

Based on the brain training literature, we predicted that training on a spatial working memory task would produce transferable gains to untrained tasks that were cognitively related ('near transfer'). As a control, we also expected significant gains in performance on a second spatial short-term memory task that was almost identical to the trained task in terms of cognitive requirements and design. Finally, we hypothesized that the same results would be found when we modified the experimental design to closely mimic that of many successful brain training studies.

2. Methods

2.1. Experiment 1

2.1.1. Participants

Participants were recruited from two research participant pools: 1) locally from the University of Western Ontario, using recruitment flyers, and 2) from Mechanical Turk (MTurk), Amazon's crowdsourcing platform. Participants recruited from MTurk who completed the task were paid $2.00 per session (which lasted approximately 30 min), and were given a $1.00 bonus for every five sessions they completed. Those recruited locally from the University of Western Ontario were paid the same amount for completing the tasks at home, but were given $10/hour to cover transportation costs if they completed the task in the lab. To be included in the analysis, participants had to 1) complete the pre-training test; 2) complete the post-training test; 3) complete at least 16 days of cognitive training with no more than 3 days between training sessions. This amounted to a minimum of approximately 10 h total training. 4) showed evidence of improved performance on the training task based on the slope of a linear fit (mean adjusted R^2). A total of 76 participants signed up for the experiment; of the 76 participants, 56 had completed the pre- and post-test, 48 of those participants had completed at least 16 days of training, and 47 had also improved on the training task. The 47 participants (26 females) between the ages of 20 and 62 (M = 32.89, SD = 8.41) who met all criteria were included in the final analysis. Our final sample size exceeds that of many other studies using different working memory tasks in context of cognitive training that show strong training effects (see Morrison and Chein, 2011). A control group (31 participants; 14 female, ages 22–53; M = 31.35, SD = 6.77) who completed the pre- and post-training 30 days apart, but did not engage in any cognitive training was also included. There were no significant differences in demographic information between the training and control groups. All participants consented to participating, and the study was approved by the Health Sciences Research Ethics Board of the University of Western Ontario.

2.1.2. Procedure

The experiment consisted of three phases: 1) pre-training, 2) training, and 3) post-training, which were completed over the course of 30 days. On the first day of the experiment (the pre-training phase), each participant completed the two test tasks, which served as a baseline measure of their ability on these tasks. The training phase started within three days of completing the pre-training phase. Within three days of finishing the training, participants completed the same test
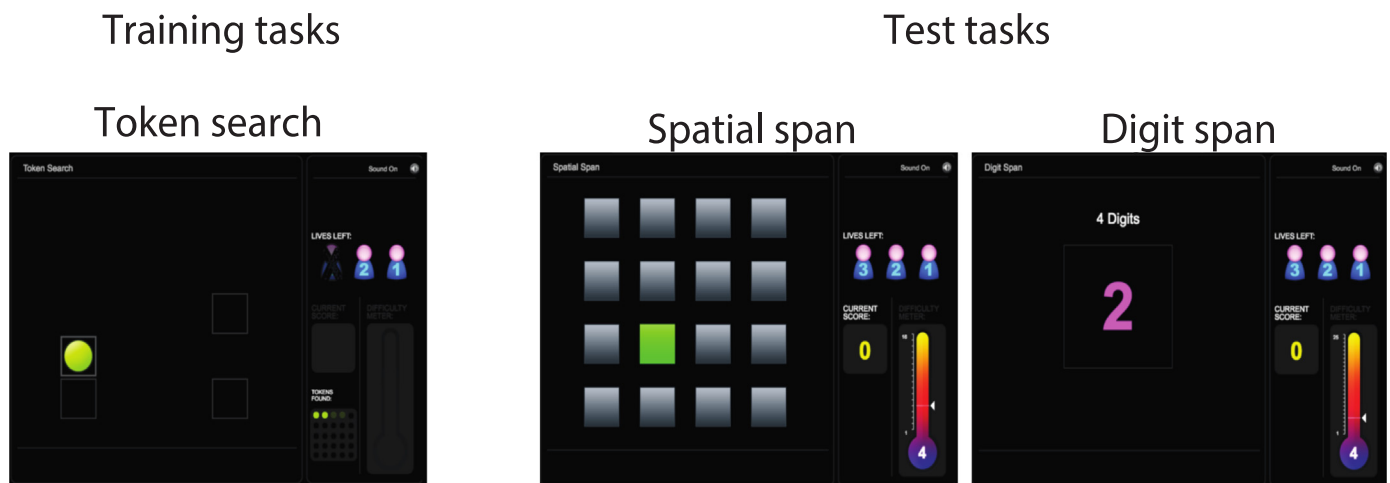
Training tasks

Token search

Test tasks

Spatial span

Digit span

Fig. 1. Training and test tasks: Screen shots of the two training tasks (on left: token search task and the two test tasks (on right: spatial span and digit span).

tasks to conclude the post-training phase. Each participant completed the test tasks three consecutive times (during both the pre- and post-training phase) to ensure an accurate and reliable measure of their ability. The control group followed the same protocol, except they did not participate in the training phase. In total, we used three tasks; 2 test tasks (henceforth referred to as 'spatial span' and 'digit span') and one training task (henceforth referred to as 'token search'). The three tasks were taken from Cambridge Brain Sciences (Owen et al., 2010), a suite of twelve online tasks that measure different aspects of cognitive function (Fig. 1).

2.1.3. Training task

In this experiment, participants trained on the token search task (Fig. 1). This task was chosen for three reasons: 1) it is a well-validated measure of short-term (working) memory (Owen et al., 1990, 1996; Hampshire et al., 2012) 2) it has an overlapping cognitive and neural profile to other short-term (working) memory tasks previously shown to have transferable gains in cognition, such as the n-back task and 3) it does not suffer from the same capacity limits that are characteristic of other short-term (working) memory tasks (Jaeggi et al., 2008; Owen et al., 2005). The task begins with a set of 4 empty squares randomly positioned on the screen. The participant's task is to find a hidden "token" by clicking on the squares to reveal their contents. Once the token has been discovered, a new token is hidden behind another square, and the participant must again find it. Once the second token has been discovered, a third will be hidden and so on until all four locations have been used to hide a token. Importantly, a token will never appear in the same square twice in the same trial; the participant must continue to search until a token has been found behind each square. If the participant clicks on an empty square twice, or clicks on a square where the token has been found previously, this is recorded as an error. The task is adaptive, which makes it ideal for brain training (Klingberg, 2010; Morrison and Chein, 2010). That is to say, the difficulty level changes based on the participant's performance; the number of squares in each trial goes down by one for every error committed, and is increased by one if a token is found under all the available squares. After three errors, the test ends. The participant's score on the task is calculated as the number of discovered tokens (or number of available squares) in the last correctly completed trial. Participants completed 10 trials per training day (30–45 min) during the training phase, amounting to an average of 19.7 training days, and approximately 13 h of training. Each individual participant's test score was computed by taking the mean across the 10 trials.

2.1.4. Test tasks

A unique feature of this study is that the test tasks (Fig. 1) were

selected based on quantifiable measures of similarity to the training task, thereby operationalizing how transfer is defined and measured. Based on a factor analysis computed by Hampshire et al. (2012) that grouped the 12 tasks that make up CBS into three factors, the assigned factor, and the corresponding factor loadings were selected to guide the choice of test tasks. The digit span task was selected as a measure of near transfer because, like the token search task, it is a short-term (working) memory task, but in a different domain (highest factor loading on the verbal component; see Hampshire et al., 2012). This verbal short-term (working) memory task requires participants to re-member a sequence of numbers that are presented in the middle of the screen one at a time. Once the entire sequence of numbers has been presented, participants must reproduce, in the exact same order, the sequence of numbers they just saw using the keyboard. On every successful trial, the number of digits in the sequence increases by one, whereas reproducing an incorrect sequence results in one less number in the following trial. After three errors the trial is over. The spatial span task was selected as the second test task, and served as a control. Spatial span is a spatial short-term memory task that loads heavily on the same component as the token search task, and is nearly identical in design, drawing on a similar set of cognitive and neural mechanisms (see Owen et al., 1990, 1996). In this task, participants are required to reproduce from memory a sequence of flashing boxes that appear randomly every 900 ms on a 4 × 4 array on the screen. The number of boxes increases, or decreases, by one for every correct and incorrect trial, respectively.

In addition to using quantifiable measures to select test tasks that have similar properties to the training task, these two tasks were chosen for another reason: performance on both tasks can be improved, even after a few trials (Bellander et al., 2011; Brehmer et al., 2012; Ericcson et al., 1980; Ericsson and Chase, 1982). This is crucial for establishing that the test tasks are sufficiently sensitive to detect any evidence of transfer.

2.1.5. Statistical analyses

Data were analysed using multiple statistical methods. Specifically, the general linear model was used to compute mixed effects, and re-peated measures ANOVAs, t-tests, and effect sizes. Bayesian statistics were also used to determine the likelihood and the strength of any effects. Since frequentist and Bayesian statistics provide complimentary perspectives to addressing the same issues (the former concerned with Type 1 and Type 2 errors, and the latter with incorporating prior probabilities that determine the likelihood a certain result falls under the null hypothesis; Lakens, 2017), both are reported. To compute Bayes Factors the statistical software JASP was used (JASP Team, 2017). The default Bayes Factor approach was utilized for model

selection using, symmetric Cauchy prior with width $\sqrt{2}/2$ which translates to a 50% confidence that the true effect will lie between -0.707 and 0.707. This evidentiary strength is expressed as a Bayes Factor (Kass and Raftery, 1995), which can be interpreted as the relative likelihood of one model versus another given the data and a certain prior expectation. A Bayes Factor of, e.g., 7, in favour of a regression model suggests that the data are seven times *more likely* under that model than an intercept only model for a given prior (for an empirical comparison of p-values and Bayes factors, see Kruschke (2013).

The data were analysed in two ways. First, data from all the trained participants were included to determine i) whether there was a significant training effect (direct transfer), ii) whether these individuals also improved on the untrained test tasks and iii) whether those who trained performed better than those in the control group (completed no brain training). To increase the likelihood of finding evidence for near transfer effects, we divided the participants into "high learners" and "low learners", based on a median split of their performance on the training task (token search). Transfer was again assessed, but using the data only from the "high learners". Note, we did not compare performance between "high learners" and "low learners" as responder analyses are limited and can introduces biases (Tidwell et al., 2014)

2.2. Experiment 2

2.2.1. Participants

Participants were recruited from MTurk, and were paid $4.00 per session (which lasted approximately 30–45 min). They were also given a $1.00 bonus for every five sessions they completed. The same inclusion criteria outlined in experiment 1 were used: participants had to complete both the pre- and post-training phase, and at least 16 days of training with no more than 3 days between training sessions (amounting to a minimum of approximately 10 h) and improve on the training task (based on the slope of a linear fit – mean adjusted R^2). A total of 24 participants (out of 64 who registered) completed the study, of which 21 participants (9 females) between the ages of 21 and 52 (M = 35.67, SD = 9.01) showed a positive benefit of training and were included in the final analysis. Our sample is similar to other studies (e.g., Dahlin et al., 2008; Jaeggi et al., 2010; Thompson et al., 2013) that have used the same task and shown strong training effects. The same control group in Experiment 1 were also included in this experiment (see details above).

2.2.2. Training task

In this experiment, participants trained on the dual n-back task (see Jaeggi et al., 2008). For this task, participants are required to monitor and respond to visually presented sequences of squares (specifically their location, much like the spatial span task) while letters are aurally presented simultaneously. Each square and letter are presented for 500 ms, during which participants must indicate whether the current stimulus (a) is a visual match (same location; left arrow key), (b) an auditory match (same letter; right arrow key), (c) both a visual and auditory match (same location and letter; both left and right arrow keys), or (d) no match (different location and letter; no key response), as the stimulus that appeared *n* trials back. Participants completed 10 blocks of the task, where each block consisted of *n* + 20 trials with randomized stimuli. The task was also adaptive; that is, the participant's performance determined the level of *n* for the following blocks of trials, where 90% correct responses in one block led to an increase of one *n* in the following block, and performance at less than 70% led to a decrease of one *n* in the following block. Performance between those thresholds did not result in a change in the *n*-value. After 10 blocks, the *n*-value of the last trial was recorded as the final score of the participants, with higher scores indicating greater short-term (working) memory capacity than lower scores. Participants completed 10 trials per training day during the training phase, amounting to an average of 18.86 training days, for a total of approximately 12.5 h of training, which is longer

than the amount of training used by Jaeggi et al. (2008), and other studies to show transferable gains after training on the dual n-back (Lilienthal et al., 2013; Salminen et al., 2016; Schweizer et al., 2011). Individual participant's test scores were computed by taking the mean across the 10 trials.

The dual n-back task was selected for three primary reasons. First, it is a measure of short-term (working) memory. Second, conceptually it is very similar to the token search task used in experiment 1 – in fact, the pattern of neural activity produced by the dual n-back and n-back tasks in general (Owen et al., 2005) overlap considerably with the pattern of activity elicited by the token search task and other short-term (working) memory tests comprising CBS (Hampshire et al., 2012). Third, the dual n-back task has been widely used in previous cognitive training studies, some of which have reported both near and far transfer effects (Jaeggi et al., 2010, 2008; Lilienthal et al., 2013; Salminen et al., 2016).

2.2.3. Test tasks

As in experiment 1, the spatial and digit span tasks were used to look for evidence of training effects (see above for explanations of the tasks). Given the considerable overlap in cognitive-neural profiles of the dual n-back and the token search task, the same logic was applied for selecting the spatial and digit span tasks for the second experiment; that is, they are commonly used tasks that measure short-term (working) memory capacity (across two domains), and are sensitive to performance improvements (Ericcson et al., 1980; Olesen et al., 2004).

2.2.4. Statistical analyses

The same analysis protocol was used as in Experiment 1: frequentist and Bayesian statistics to evaluate the reliability, strength and likelihood of any transfer effects and differences with the control group (see statistical analysis in Experiment 1 for more information).

3. Results

3.1. Experiment 1

3.1.1. Training

Using the data from all the participants who met the criteria for data inclusion (see methods), a paired samples *t*-test and a Bayesian paired sample *t*-test were computed to determine whether participants improved on the training task (token search task). Performance on the last day of training (after at least 16 days) was significantly better than performance on the first day of training (Fig. 2a&b; $p < 0.001$; Cohen's d = 0.63; Bayes factor (BF_{10} = 385.09); r = 0.985), with participants improving on average by 18%. Not surprisingly, even stronger training effects were found when performance of "high learners" was compared on the first and last day of training (Fig. 4a&b; $p < 0.001$, Cohen's d = 1.18; BF_{10} = 2367); maximum performance was reached only after 16 days, with a mean improvement of approximately 20%.

3.1.2. Transfer

To examine whether training on the token search task resulted in improved performance on the two test tasks, and whether that improvement was greater than the control group, we computed a mixed effects ANOVA, and Bayesian ANOVA, with group as the between-subjects factor (training on the token search task vs. control), and task (digit span vs. spatial span) testing day (pre-test vs. post-test) as the within-subjects factor. With the exception of a main effect of task ($F_{(1,298)}$ = 116.7; $p < 3.5E\text{-}20$; η^2 = 0.281; BF_{10} = 1.15E21), the analysis revealed no other significant effects. This included, no main effects of testing day ($F_{(1,298)}$ = 0.2; p = 0.72; η^2 = 3.36E-4; BF_{10} = 0.13; 1-β = 0.061), or group ($F_{(1,298)}$ = 1.81; p = 0.18; η^2 = 0.006; BF_{10} = 0.29), nor was there a significant task by testing day interaction ($F_{(1,298)}$ = 0.005; p = 0.93; η^2 = 1.62E-5; BF_{10} = 0.167), group by task interaction ($F_{(1,298)}$ = 0.04; p = 0.85; η^2 = 1.19E-4; BF_{10} = 0.168), and testing day by group ($F_{(1,298)}$ = 1.39; p = 0.24; η^2 = 0.005; BF_{10} =
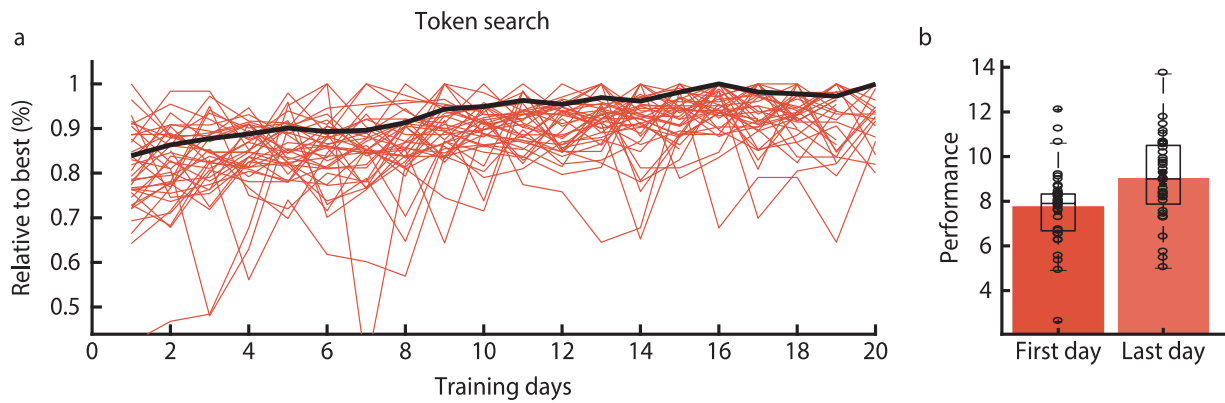
Fig. 2. Performance on the Token search task throughout training phase. A. Red lines represent changes in performance for each participant relative to their best score across the 20 days of training (group level performance is marked by the black line). B. Performance on the Token search task significantly improved from the first to last day of training; Bars plots represent mean performance, and box and whiskers plots depicting median score enclosed by from 25th to 75th quartiles; whiskers extend to maximum and minimum values not considered outliers. Individual scores are shown in black circles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

0.28).

To ensure we did not overlook subtle transfer effects for the group that trained on the token search task, we ran a repeated measures ANOVA, and a Bayesian repeated measures (task vs. testing day) were computed to compare changes in behavioural performance on the digit span and spatial span tasks before and after training. The repeated measures ANOVA revealed no main effect of testing day ($F_{(1,46)} = 2.24$; $p = 0.14$; $\eta^2 = 0.05$; $1\text{-}\beta = 0.063$) and no task by testing day interaction ($F_{(1,46)} = 0.002$; $p = 0.97$; $\eta^2 = 3.37\text{E-}5$). A main effect of test was observed ($F_{(1,46)} = 61.5$; $p < 0.001$) reflecting the fact that the two tests use different rating scales. The Bayesian repeated measures ANOVA revealed substantial support for the null finding with a Bayes Factor ($BF_{10} = 0.21$) for the main effect of testing day and $BF_{10} = 0.22$ for the task by testing interaction. Together these results indicate that training on the token search task did not transfer to performance improvements on either the digit span or the spatial span tasks (Fig. 3a& b).

The same pattern of results was found for the "high learners" (those who benefited the most from training; Fig. 4a&b); not only did their performance on the digit span not change from the pre-training to the post-training, but performance on the spatial span was also not significantly improved (Fig. 4b&c; main effect of testing day: $F_{(1,22)} = 1.86$; $p = 0.19$; $\eta^2 = 0.08$; $1\text{-}\beta = 0.066$; $BF_{10} = 0.28$; and task by testing day interaction; $F_{(1,22)} = 0.15$; $p = 0.702$; $\eta^2 = 0.007$; $BF_{10} = 0.31$). Pairwise comparisons between pre-training and post-training scores on the spatial and digit span tasks were not significant for any number of training sessions greater than 16.

3.2. Experiment 2

3.2.1. Training and transfer

In the second experiment, cognitive training resulted in significant improvements on the dual N-back on the last day when compared to the first day, (Fig. 5a&b; $p < 0.001$; Cohen's d = 1.41; Bayes Factor ($BF_{10} = 7374$) with improvements exceeding 60%. Despite extensive training, (exceeding that of many previous studies using the dual N-back task), the results of a mixed effects ANOVA, and a Bayesian ANOVA (with group as the between subjects factor (training on the token search task vs. control), and task (digit span vs. spatial span) testing day (pre-test vs. post-test) as the within subjects factor) revealed no transferable benefit to performance on the digit or spatial span tasks, (main effect of testing day: $F_{(1,200)} = 0.02$; $p = 0.89$; $\eta^2 = 6.33\text{E-}5$; $BF_{10} = 0.15$; $1\text{-}\beta = 0.05$; task by testing day interaction: $F_{(1,200)} = 0.032$; $p = 0.86$; $\eta^2 = 1.1\text{E-}4$; $BF_{10} = 0.21$), Moreover, we failed to find differences between the training and the control groups on either

testing task (main effect of group: $F_{(1,200)} = 4.8\text{E-}4$; $p = 0.98$; $\eta^2 = 1.69\text{E-}6$; $BF_{10} = 0.15$; task by group: $F_{(1,200)} = 0.4$; $p = 0.53$; $\eta^2 = 0.001$; $BF_{10} = 0.23$; testing day by group: $F_{(1,200)} = 1.82$; $p = 0.53$; $\eta^2 = 0.006$; $BF_{10} = 0.35$. Again, to ensure transfer effects within the training group were not overlooked, we conducted a repeated measures ANOVA, and a Bayesian ANOVA and found the same set of results (Fig. 5c&d; main effect testing day $F_{(1,20)} = 3.53$; $p = 0.08$; $\eta^2 = 0.15$; $1\text{-}\beta = 0.098$; $BF_{10} = 0.32$; task by testing day interaction; $F_{(1,20)} = 0.23$; $p = 0.641$; $\eta^2 = 0.01$; $BF_{10} = 0.318$). Including additional training days did not produce significant changes in performance on either test task.

4. Discussion

The goal of this study was to investigate whether targeted brain training protocols produce generalizable improvements to cognition. Across two experiments, participants trained on either the token search task (Exp. 1) or the dual n-back task (Exp. 2), and were tested on the spatial and digit span tasks. These tasks were chosen strategically using quantifiable measures, in order to constrain how transfer is defined, and maximized the likelihood of finding evidence for the benefits of brain training. If the principles of brain training reflect the underlying properties of cognition, this procedure should have produced evidence in support of that.

The results from experiment 1 indicated that participants who trained on the token search task performed significantly better at the end of the training phase compared to the pre-training phase, improving by approximately 18%. This result is important because it demonstrates for the first time that performance on the CBS token search task can be improved with training, establishing that it is a viable task for brain training programs. Despite this improvement, however, we found no evidence that it resulted in transferable gains in performance on the untrained test tasks; participants did not improve on either the digit span or the spatial span tasks despite their high degree of similarity to the training task. The lack of improvement was particularly surprising in the case of spatial span, which is conceptually almost identical in design and implementation to the training task (both require participants to remember the location of boxes on the screen), as well as in terms of the cognitive mechanisms that are required (Owen et al., 1990), and the underlying neural structures that are recruited (Owen et al., 1996). A similar pattern of results was found in experiment 2. While participants significantly improved on the dual n-back, by a margin comparable to other training studies using the dual n-back (Heinzel et al., 2014; Jaeggi et al., 2010, 2008; Lilienthal et al., 2013), their performance on the digit span and spatial span tasks did not
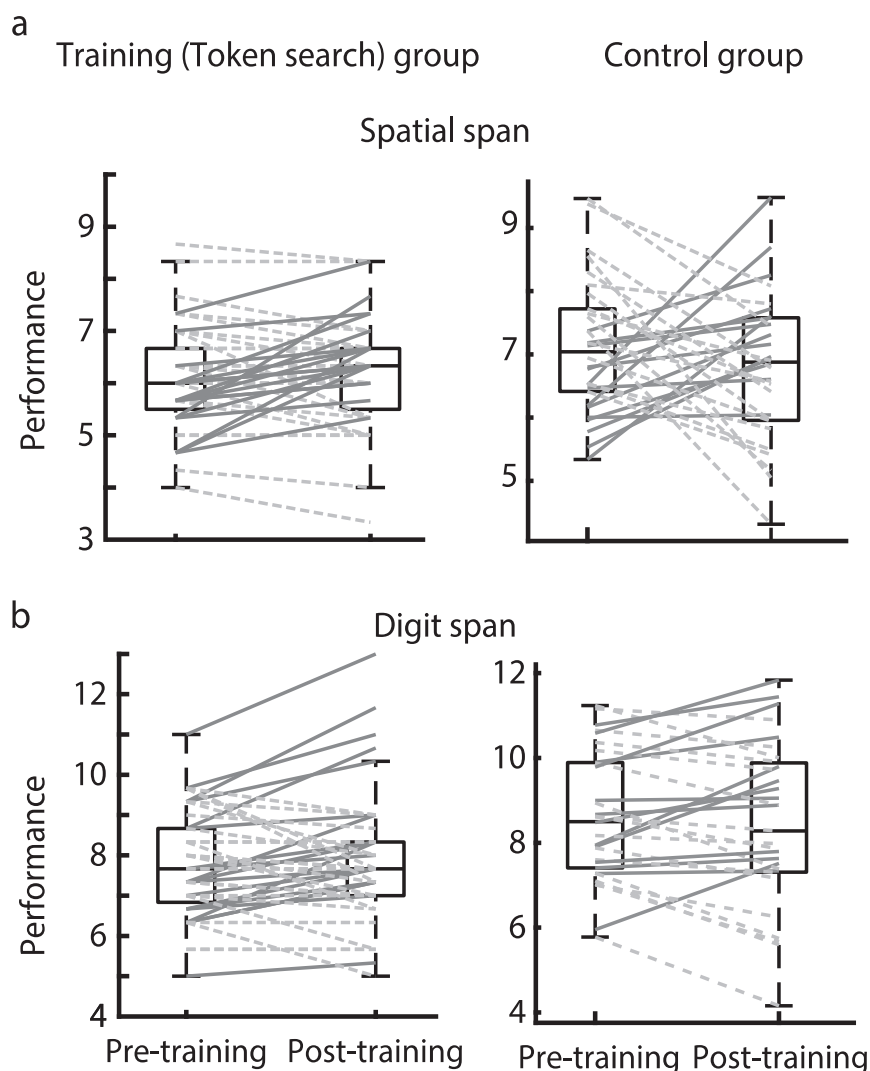
a

Training (Token search) group   Control group

Spatial span

b

Digit span

Fig. 3. Performance on the test tasks. A. No difference in performance on the spatial span task at pre-training and post-training, for either the training group (left column) or the control group (right column) We also found no difference in performance between the two groups. Box and whisker plot depict median performance (horizontal black line) the edge of the box marks the 25th and 75th quartiles. The whiskers indicate the highest and lowest values not considered outliers. Grey lines represent changes in performance for each participant; solid lines indicate improved performance and dashed lines indicated a decrease in performance. B. The same set of results was found for the digit span task; the same plots were used as in figure a but for performance on the digit span.

change after training. Taken together, the results of both experiments indicate that performance on the digit span and spatial span tasks was nearly identical independent of training program – that is, whether participants trained on the token search task (even for "high learners") or the dual n-back task. In fact, all participants who completed the training, scored similarly to a control group who completed both test tasks, but did not engage in any form of training.

Is it possible that these results can be explained by factors unrelated to brain training that may have interfered with our ability to detect transferable improvements in cognition? For example, were the wrong training tasks used? This is unlikely because the dual n-back task has been used widely across a number of brain training studies, some of which have found evidence to support generalizable improvements to short-term memory following training (Jaeggi et al., 2010, 2008; Lilienthal et al., 2013). Moreover, the token search task shares many of the same properties as the dual n-back; they activate a similar network of brain regions and rely on similar cognitive processes. If short-term memory-based brain training reliably produced global benefits to cognition, those effects should have been replicated in these experiments, and they should have extended to the token search task.

A second possible reason why transferable improvement to untrained short-term memory tasks were not observed is because the test tasks used were not sufficiently sensitive to the effects of training. This is also unlikely for three reasons: first, the spatial span (Bellander et al., 2011; Chein and Morrison, 2010; Olesen et al., 2004) and the digit span

tasks (Ericcson et al., 1980; Ericsson and Chase, 1982) have been shown to be sensitive to improvement, even after subtle changes in cognitive function due to practice (Ericcson et al., 1980), disease, or pharmacological intervention (Owen et al., 2010). Second, both the digit span and the spatial span tasks (or variants of them) are commonly used as test tasks in studies that test for the degree of transfer after training, and in some cases performance on both of these tasks improved (Caeyenberghs et al., 2016; Jaeggi et al., 2008; Lilienthal et al., 2013). Third, the test tasks were selected because they are quantifiably very similar to the training tasks, based on overlapping cognitive, and neural profiles (Hampshire et al., 2012). By doing so, it was possible to ensure that the cognitive system that is assumed to be enhanced by training (in this case short-term memory), is also the primary cognitive system required to complete the test tasks. Indeed, Klingberg et al. (2010) have argued that training on short-term memory produces neural plasticity in brain regions, and networks supporting short-term memory processes. If this principle accurately characterizes the properties underlying brain training, the experimental design used here should have detected the behavioural manifestations of these changes, resulting in generalizable improvement across similar tasks.

Could it be that the training protocol used in the current experiments was too short, and participants did not receive enough training (and therefore, improve sufficiently on the training task), to improve their scores on untrained tasks? Again, this factor does not provide an adequate explanation for the results observed. Our participants trained
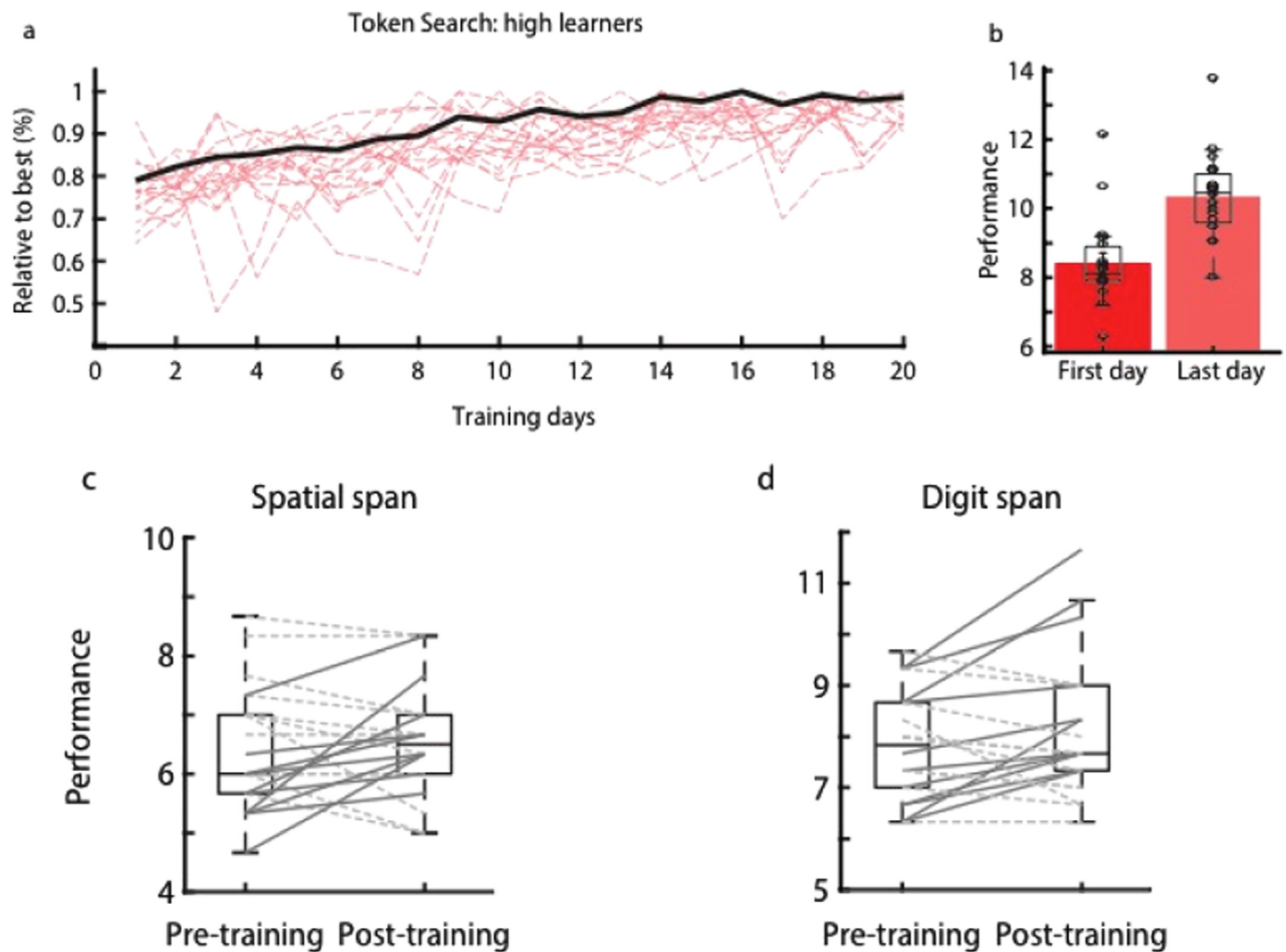
Fig. 4. Performance on the Token search and test tasks (Spatial and Digit span). A. Change in performance for "high learners" (N = 23) relative to their best score, throughout the training phase (dashed red line), with the black line representing group level changes in performance. B. Significant improvement on the Token search task on the first and last day of training; mean performance represented by bar plots (± standard error) and distribution of individual scores (black circles) are superimposed along with box and whiskers plots (median score with the edges of the box marking 25th to 75th quartiles and whiskers extending to maximum and minimum values not considered outliers). C,D. Despite significant improvements to the Token search task during training, this did not transfer to performance on the Spatial span task or the Digit span task; box and whiskers plots superimposed with changes in performance for each participant (solid line representing those who improved and dashed line representing those who did not). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

on both the token search task and the dual n-back task for approximately 13, and 12.5 h, respectively, an amount consistent with most brain training studies (Anguera et al., 2012; Heinzel et al., 2014; Redick et al., 2013; Thompson et al., 2013) and longer than many studies that have reported transfer effects (Jaeggi et al., 2010, 2008; Lilienthal et al., 2013), which featured prominently in our design to replicate those findings. However, if brain train training results in global gains in short-term memory, showing significant improvements on the training task should be the only requirement, independent of training duration (Jaeggi et al., 2010). In these experiments, very large improvements were observed on both training tasks. Indeed, the degree to which participants improved on the token search task is comparable to the amount participants improved on short-term memory tasks used in other brain training studies (for example, (Holmes et al., 2009; Morrison and Chein, 2010), and participants improved on the dual n-back to a level at least equivalent to many other studies that used it as a training task, importantly, including studies that found transferable gains after training (Anguera et al., 2012; Jaeggi et al., 2010, 2008; Salminen et al., 2016; Thompson et al., 2013). In fact, the duration of the training protocol used here may have been longer than necessary;

participants reached maximum and stable improvement on both training tasks after the 16th day, suggesting that more training would not have significantly improved performance on the training task, and would therefore likely have no additional benefit to performance on the test tasks either. Even those who benefited the most from the training did not improve on the untrained short-term memory tasks. Finally, performance on the test tasks by the control group did not differ from performance on the same tasks after training on either the token search or the dual n-back tasks.

Perhaps our null findings are due to potential biases associated with on-line data collection. This is unlikely to affect our results. Accuracy of online data has been shown to be reliable and valid (Morrison et al., 2015; Rosa et al., 2014; Ruano et al., 2016), and data obtained from online platforms, such as Mechanical-Turk, are not only of high-quality (Crump et al., 2013), but have been used to replicate various psychological findings (Buhrmester et al., 2011). Beyond the general utility of online data collection, the CBS platform used in the current study has also been used successfully in previous large-scale studies (Hampshire et al., 2012; Owen et al., 2010). Moreover, we found no difference in performance on the token search task (Experiment 1) between
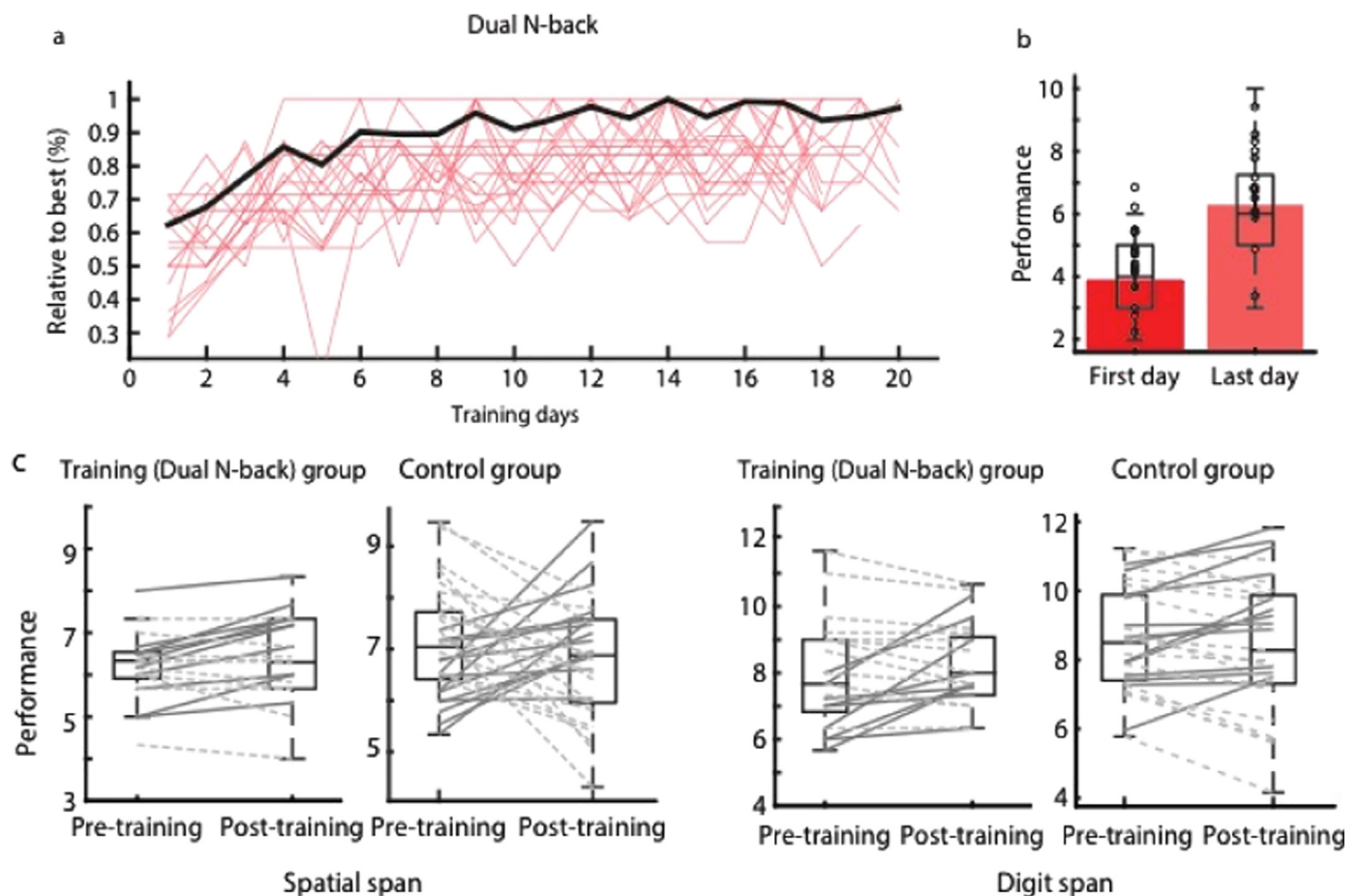
Fig. 5. Experiment 2 results: Performance on the training (Dual n-back) and test tasks (Spatial and Digit span). A. Red line represents improvement (relative to best score) during the training phase for each participant; black line reflects group level change in performance (approximately 40% improvement). B. Participants (black circles) improved significantly on the Dual n-back task from the last day relative to the first day; mean performance depicted by bar plots and distribution of scores represented by box and whisker plot marking median performance enclosed by 25th and 75th quartiles and whiskers extending to maximum and minimum scores. C,D. Improvement on the dual n-back task did not generalize to performance on the Spatial span task) or the Digit span task, and performance on both test tasks on either testing day were no different than that of the control group; bar plots (± standard error) illustrate mean performance before and after training. Change in performance on both test tasks are shown in grey lines (solid dark grey for those who improved and dashed light grey for those who did not) plus box and whisker plots depicting median performance along with maximum and minimum scores. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

participants who were recruited online (via Mechanical-Turk) and those recruited at the University (see Supplementary Figure 1). Therefore, while there are a number of potential limitations associated with collecting data using various online platforms (Buhrmester et al., 2011), our lack of evidence in support of brain training cannot be accounted for based on this method of data collection. The findings reported here provide compelling evidence that targeted brain training does not produce generalizable improvements on untrained short-term memory tasks in healthy participants. This was the case despite all efforts to maximize the likelihood of finding evidence to support the purported benefits of brain training; two different short-term memory tasks were selected, with overlapping cognitive and neural profiles (Hampshire et al., 2012; Owen et al., 2005). Quantifiable measures were employed to select tests tasks that were more similar to the training task than is commonly the case in other studies. Together, the results suggest that brain training protocols that focus on increasing the capacity of short-term memory do not yield generalizable improvements to cognition, regardless of the specific training task employed.

Acknowledgements

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.neuropsychologia.2018.07.013.

References

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

Anguera, Joaquin A., Bernard, J.A., Jaeggi, S.M., Buschkuehl, M., Benson, B.L., Jennett, S., Seidler, R.D., 2012. The effects of working memory resource depletion and training on sensorimotor adaptation. Behav. Brain Res. 228 (1), 107–115. https://doi.org/10.1016/j.bbr.2011.11.040.

Au, J., Buschkuehl, M., Duncan, G.J., Jaeggi, S.M., 2016. There is no convincing evidence that working memory training is NOT effective: a reply to Melby-Lervåg and Hulme (2015). Psychon. Bull. Rev. 23 (1), 331–337. https://doi.org/10.3758/s13423-015-0967-4.

Au, J., Sheehan, E., Tsai, N., Duncan, G.J., Buschkuehl, M., Jaeggi, S.M., 2014. Improving fluid intelligence with training on working memory: a meta-analysis. Psychon. Bull. Rev. 22 (2), 366–377. https://doi.org/10.3758/s13423-014-0699-x.

Baddeley, A., 1992. Working memory. Science 255 (5044), 556–559.

Basak, C., Boot, W.R., Voss, M.W., Kramer, A.F., 2008. Can training in a real-time strategy videogame attenuate cognitive decline in older adults? Psychol. Aging 23 (4), 765–777. https://doi.org/10.1037/a0013494.

Bellander, M., Brehmer, Y., Westerberg, H., Karlsson, S., Fürth, D., Bergman, O., Bäckman, L., 2011. Preliminary evidence that allelic variation in the LMX1A gene

influences training-related working memory improvement. Neuropsychologia 49 (7), 1938–1942. https://doi.org/10.1016/j.neuropsychologia.2011.03.021.

Bergman-Nutley, S., Klingberg, T., 2014. Effect of working memory training on working memory, arithmetic and following instructions. Psychol. Res. 78 (6), 869–877. https://doi.org/10.1007/s00426-014-0614-0.

Bleiberg, J., Cernich, A.N., Cameron, K., Sun, W., Peck, K., Ecklund, J., Warden, D.L., 2004. Duration of cognitive impairment after sports concussion. Neurosurgery 54 (5), 1073–1080. https://doi.org/10.1227/01.NEU.0000118820.33396.6A.

Brehmer, Y., Westerberg, H., Bäckman, L., 2012. Working-memory training in younger and older adults: training gains, transfer, and maintenance. Front. Hum. Neurosci. 6. https://doi.org/10.3389/fnhum.2012.00063.

Buhrmester, M., Kwang, T., Gosling, S.D., 2011. Amazon's mechanical turk a new source of inexpensive, yet high-quality, data? Perspect. Psychol. Sci. 6 (1), 3–5. https://doi.org/10.1177/1745691610393980.

Caeyenberghs, K., Metzler-Baddeley, C., Foley, S., Jones, D.K., 2016. Dynamics of the human structural connectome underlying working memory training. J. Neurosci. 36 (14), 4056–4066. https://doi.org/10.1523/JNEUROSCI.1973-15.2016.

Calvin, C.M., Deary, I.J., Fenton, C., Roberts, B.A., Der, G., Leckenby, N., Batty, G.D., 2011. Intelligence in youth and all-cause-mortality: systematic review with meta-analysis. Int. J. Epidemiol. 40 (3), 626–644. https://doi.org/10.1093/ije/dyq190.

Chein, J.M., Morrison, A.B., 2010. Expanding the mind's workspace: training and transfer effects with a complex working memory span task. Psychon. Bull. Rev. 17 (2), 193–199. https://doi.org/10.3758/PBR.17.2.193.

Crump, M.J.C., McDonnell, J.V., Gureckis, T.M., 2013. Evaluating Amazon's mechanical turk as a tool for experimental behavioral research. PLoS One 8 (3), e57410. https://doi.org/10.1371/journal.pone.0057410.

Dahlin, E., Neely, A.S., Larsson, A., Bäckman, L., Nyberg, L., 2008. Transfer of learning after updating training mediated by the striatum. Science 320 (5882), 1510–1512. https://doi.org/10.1126/science.1155466.

Dahlin, K.I.E., 2011. Effects of working memory training on reading in children with special needs. Read. Writ. 24 (4), 479–491. https://doi.org/10.1007/s11145-010-9238-y.

Enge, S., Behnke, A., Fleischhauer, M., Küttler, L., Kliegel, M., Strobel, A., 2014. No evidence for true training and transfer effects after inhibitory control training in young healthy adults. J. Exp. Psychol. Learn. Mem. Cogn. 40 (4), 987–1001. https://doi.org/10.1037/a0036165.

Engle, R.W., Tuholski, S.W., Laughlin, J.E., Conway, A.R., 1999. Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. J. Exp. Psychol. Gen. 128 (3), 309–331.

Ericcson, K.A., Chase, W.G., Faloon, S., 1980. Acquisition of a memory skill. Science 208 (4448), 1181–1182. https://doi.org/10.1126/science.7375390.

Ericsson, K.A., Chase, W.G., 1982. Exceptional memory: extraordinary feats of memory can be matched or surpassed by people with average memories that have been improved by training. Am. Sci. 70 (6), 607–615.

Foroughi, C.K., Monfort, S.S., Paczynski, M., McKnight, P.E., Greenwood, P.M., 2016. Placebo effects in cognitive training. Proc. Natl. Acad. Sci. USA 113 (27), 7470–7474. https://doi.org/10.1073/pnas.1601243113.

Hampshire, A., Highfield, R.R., Parkin, B.L., Owen, A.M., 2012. Fractionating human intelligence. Neuron 76 (6), 1225–1237. https://doi.org/10.1016/j.neuron.2012.06.022.

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 in young and older adults. Aging Neuropsychol. Cogn. 21 (2), 146–173. https://doi.org/10.1080/13825585.2013.790338.

Holmes, J., Gathercole, S.E., Dunning, D.L., 2009. Adaptive training leads to sustained enhancement of poor working memory in children. Dev. Sci. 12 (4), F9–F15. https://doi.org/10.1111/j.1467-7687.2009.00848.x.

Jaeggi, S.M., Buschkuehl, M., Jonides, J., Perrig, W.J., 2008. Improving fluid intelligence with training on working memory. Proc. Natl. Acad. Sci. USA. 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 38 (6), 625–635. https://doi.org/10.1016/j.intell.2010.09.001.

JASP Team (2018). JASP (Version 0.9) [Computer software].

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. https://doi.org/10.1523/JNEUROSCI.2832-16.2017.

Kane, M.J., Engle, R.W., 2002. The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual-differences perspective. Psychon. Bull. Rev. 9 (4), 637–671.

Kane, M.J., Hambrick, D.Z., Tuholski, S.W., Wilhelm, O., Payne, T.W., Engle, R., 2004. The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning. J. Exp. Psychol.: Gen. 133 (2), 189–217. https://doi.org/10.1037/0096-3445.133.2.189.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90 (430), 773–795. https://doi.org/10.1080/01621459.1995.10476572.

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

Klingberg, T., Fernell, E., Olesen, P.J., Johnson, M., Gustafsson, P., Dahlström, K., Westerberg, H., 2005. Computerized training of working memory in children with ADHD-a randomized, controlled trial. J. Am. Acad. Child Adolesc. Psychiatry 44 (2), 177–186. https://doi.org/10.1097/00004583-200502000-00010.

Kruschke, J.K., 2013. Bayesian estimation supersedes the *t*-test. J. Exp. Psychol.: General. 142 (2), 573–603. ⟨http://dx.doi.org.proxy1.lib.uwo.ca/10.1037/a0029146⟩.

Lakens, D., 2017. Equivalence Tests: A Practical Primer for t Tests, Correlations, and Meta-Analyses. Soc. Psychol. Personal. Sci. 8 (4), 355–362. https://doi.org/10.1177/1948550617697177.

Lee, H., Boot, W.R., Basak, C., Voss, M.W., Prakash, R.S., Neider, M., Kramer, A.F., 2012. Performance gains from directed training do not transfer to untrained tasks. Acta Psychol. 139 (1), 146–158. https://doi.org/10.1016/j.actpsy.2011.11.003.

Li, S.-C., Schmiedek, F., Huxhold, O., Röcke, C., Smith, J., Lindenberger, U., 2008. Working memory plasticity in old age: practice gain, transfer, and maintenance. Psychol. Aging 23 (4), 731–742. https://doi.org/10.1037/a0014343.

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

Marinus, J., Visser, M., Verwey, N.A., Verhey, F.R.J., Middelkoop, H.A.M., Stiggelbout, A.M., van Hilten, J.J., 2003. Assessment of cognition in Parkinson's disease. Neurology 61 (9), 1222–1228. https://doi.org/10.1212/01.WNL.0000091864.39702.1C.

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. https://doi.org/10.1037/a0017619.

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. Perspect. Psychol. Sci. 11 (4), 512–534. https://doi.org/10.1177/1745691616635612.

Morrison, A.B., Chein, J.M., 2010. Does working memory training work? The promise and challenges of enhancing cognition by training working memory. Psychon. Bull. Rev. 18 (1), 46–60. https://doi.org/10.3758/s13423-010-0034-0.

Morrison, G.E., Simone, C.M., Ng, N.F., Hardy, J.L., 2015. Reliability and validity of the neurocognitive performance test, a web-based neuropsychological assessment. Front. Psychol. 6, 1652. https://doi.org/10.3389/fpsyg.2015.01652.

Muller, M., Grobbee, D.E., Aleman, A., Bots, M., van der Schouw, Y.T., 2007. Cardiovascular disease and cognitive performance in middle-aged and elderly men. Atherosclerosis 190 (1), 143–149. https://doi.org/10.1016/j.atherosclerosis.2006.01.005.

Olesen, P.J., Westerberg, H., Klingberg, T., 2004. Increased prefrontal and parietal activity after training of working memory. Nat. Neurosci. 7 (1), 75–79. https://doi.org/10.1038/nn1165.

Owen, A.M., Hampshire, A., Grahn, J.A., Stenton, R., Dajani, S., Burns, A.S., Ballard, C.G., 2010. Putting brain training to the test. Nature 465 (7299), 775–778.

Owen, A.M., McMillan, K.M., Laird, A.R., Bullmore, E., 2005. N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. Hum. Brain Mapp. 25 (1), 46–59. https://doi.org/10.1002/hbm.20131.

Owen, A.M., Doyon, J., Petrides, M., Evans, A.C., 1996. Planning and spatial working memory examined with positron emission tomography (PET). Eur. J. Neurosci. 8, 353–364.

Owen, A.M., Downes, J.D., Sahakian, B.J., Polkey, C.E., Robbins, T.W., 1990. Planning and spatial working memory following frontal lobe lesions in Man. Neuropsychologia 28 (10), 10211034.

Redick, T.S., Shipstead, Z., Harrison, T.L., Hicks, K.L., Fried, D.E., Hambrick, D.Z., Engle, R.W., 2013. No evidence of intelligence improvement after working memory training: a randomized, placebo-controlled study. J. Exp. Psychol. Gen. 142 (2), 359–379. https://doi.org/10.1037/a0029082.

Redick, T.S., Shipstead, Z., Meier, M.E., Montroy, J.J., Hicks, K.L., Unsworth, N., Engle, R.W., 2016. Cognitive predictors of a common multitasking ability: contributions from working memory, attention control, and fluid intelligence. J. Exp. Psychol.: General. 145 (11), 1473–1492. ⟨http://dx.doi.org.proxy1.lib.uwo.ca/10.1037/xge0000219⟩.

Ritchie, S.J., Bates, T.C., Der, G., Starr, J.M., Deary, I.J., 2013. Education is associated with higher later life IQ scores, but not with faster cognitive processing speed. Psychol. Aging 28 (2), 515–521. https://doi.org/10.1037/a0030820.

Rosa, E.D., Hannigan, C., Brennan, S., Reilly, R., Rapčan, V., Robertson, I.H., 2014. Reliability and validity of the automatic cognitive assessment delivery (ACAD). Front. Aging Neurosci. 6. https://doi.org/10.3389/fnagi.2014.00034.

Ruano, L., Sousa, A., Severo, M., Alves, I., Colunas, M., Barreto, R., Cruz, V.T., 2016. Development of a self-administered web-based test for longitudinal cognitive assessment. Sci. Rep. 6, srep19114. https://doi.org/10.1038/srep19114.

Salminen, T., Kühn, S., Frensch, P.A., Schubert, T., 2016. Transfer after dual n-back training depends on striatal activation change. J. Neurosci. 36 (39), 10198–10213. https://doi.org/10.1523/JNEUROSCI.2305-15.2016.

Schweizer, S., Hampshire, A., Dalgleish, T., 2011. Extending brain-training to the affective domain: increasing cognitive and affective executive control through emotional working memory training. PLoS One 6 (9), e24372. https://doi.org/10.1371/journal.pone.0024372.

Simon, A.J., Skinner, S.N., Ziegler, D.A., 2016. Training working memory: anatomy matters. J. Neurosci. 36 (30), 7805–7806. https://doi.org/10.1523/JNEUROSCI.1513-16.2016.

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? Psychol. Sci. Public Interest 17 (3), 103–186. https://doi.org/10.1177/1529100616661983.

Swanson, H.L., Jerman, O., 2006. Math Disabilities: A Selective Meta-Analysis of the Literature. Rev. Educ. Res. 76, 249–274. https://doi.org/10.3102/00346543076002249.

Süß, H.-M., Oberauer, K., Wittmann, W.W., Wilhelm, O., Schulze, R., 2002. Working-memory capacity explains reasoning ability—and a little bit more. Intelligence 30 (3), 261–288. https://doi.org/10.1016/S0160-2896(01)00100-3.

Talavage, T.M., Nauman, E.A., Breedlove, E.L., Yoruk, U., Dye, A.E., Morigaki, K.E.,

Leverenz, L.J., 2014. Functionally-detected cognitive impairment in high school football players without clinically-diagnosed concussion. J. Neurotrauma 31 (4), 327–338. https://doi.org/10.1089/neu.2010.1512.

Thompson, T.W., Waskom, M.L., Garel, K.-L.A., Cardenas-Iniguez, C., Reynolds, G.O., Winter, R., Gabrieli, J.D.E., 2013. Failure of working memory training to enhance cognition or intelligence. PLoS One 8 (5), e63614. https://doi.org/10.1371/journal.pone.0063614.

Tidwell, J.W., Dougherty, M.R., Chrabaszcz, J.R., Thomas, R.P., Mendoza, J.L., 2014. What counts as evidence for working memory training? Problems with correlated gains and dichotomization. Psychon. Bull. Rev. 21 (3), 620–628. https://doi.org/10.3758/s13423-013-0560-7.

Tulbure, B.T., Siberescu, I., 2013. Cognitive training enhances working memory capacity in healthy adults. A pilot study. Procedia - Soc. Behav. Sci. 78, 175–179. https://doi.org/10.1016/j.sbspro.2013.04.274.

Westerberg, H., Jacobaeus, H., Hirvikoski, T., Clevberger, P., Östensson, M.-L., Bartfai, A., Klingberg, T., 2007. Computerized working memory training after stroke–a pilot study. Brain Inj. 21 (1), 21–29. https://doi.org/10.1080/02699050601148726.