Formation of abstract task representations: Exploring dosage and mechanisms of working memory training effects

Nitzan Shahar[1], Maayan Pereg[1], Andrei R. Teodorescu[2], Rani Moran[34], Anat Karmon-Presser[1] & Nachshon Meiran[1]

[1] Psychology department, Ben-Gurion University of the Negev, P.O.B. 653 Beer-Sheva 8410501, Israel

[2] Psychology department, University of Haifa, 199 Aba Khoushy Ave. Mount Carmel, Haifa, Israel

[3] Max Planck UCL Centre for Computational Psychiatry and Ageing Research, University College London, 10-12 Russell Square, London WC1B 5EH, UK

[4] Wellcome Centre for Human Neuroimaging, University College London, London WC1N 3BG, United Kingdom

Author Note

Correspondence concerning this article should be sent to the first author Nitzan Shahar, Department of Psychology, Ben-Gurion University of the Negev, Beer-Sheva, Israel 84105, shahar.nitzan@gmail.com

Keywords: Cognitive training, Working memory, Executive functions, Evidence accumulation

Abstract

Working memory is strongly involved in human reasoning, abstract thinking and decision making. Past studies have shown that working memory training generalizes to untrained working memory tasks with similar structure (near-transfer effect). Here, we focused on two questions: First, we ask how much training might be required in order to find a reliable near-transfer effect? Second, we ask which choice-mechanism might underlie training benefits? Participants were allocated to one of three groups: working-memory training (combined set-shifting and N-back task), active-control (visual search) and no-contact control. During pre/post testing, all participants completed tests tapping procedural and declarative working memory as well as reasoning. We found improved performance only in the procedural working-memory transfer tasks, a transfer task that shared a similar structure to that of the training task. Intermediate testing throughout the training period suggest that this effect emerged as soon as after 2 training sessions. We applied evidence accumulation modeling to investigate the choice process responsible for these near-transfer effect and found that trained participants, compared with active-control had quicker retrieval of the action rules, and more efficient classification of the target. We hypothesize that participants might be able to form abstract representations of the task procedure (i.e., stimulus-response rules) that can then be applied to novel stimuli and responses.

1. Introduction

Working memory is an attentional-cognitive control system that is considered to play a major role in goal-directed behavior and decision-making (Kane & Engle, 2002; Oberauer, 2009). It allows the agent to hold, update and manipulate relevant information in mind, while resisting interference from irrelevant information (Carruthers, 2013; Kane & Engle, 2002; Oberauer, 2009). Working memory is involved in abstract thinking, planning and reasoning (Baddeley, 2003; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002). Moreover, working memory deficits were reported in clinical conditions including; attention-deficit disorders (Andreou et al., 2007; Shahar, Teodorescu, Karmon-Presser, Anholt, & Meiran, 2016) and low intelligence (Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007; Wilhelm & Oberauer, 2006). Understanding the underlying mechanisms might therefore be of value to those conditions, especially given the potential to improve working memory via computerized training, for example.

Computerized working memory training has gained much interest over the last decade, with many studies exploring whether training can be used as a remedy for psychopathology and/or enhance human performance in healthy individuals (Klingberg, 2010; Melby-Lervåg & Hulme, 2013). An early training theory pertained a muscle-like assumption, claiming that by repeatedly loading a certain cognitive process, one might enhance the overall resources dedicated to that process. Under this assumption, improvement in a working memory demanding training task should generalize (at least in part) to other situations where working memory load is also demanding. This should be true when the amount of shared features between the training and transfer task is high (i.e., near transfer), or even when it is low (i.e., far transfer) (Constantinidis & Klingberg, 2016; Lindenberger, Wenger, & Lövdén, 2017; Melby-Lervåg, Redick, & Hulme, 2016). Early optimistic reports in support of the muscle-like assumption (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008) were later shadowed by studies claiming that far-transfer findings are mostly due to the type of control group used (Redick et al., 2013; Shipstead, Redick, & Engle, 2012). A current meta-analysis found a close to zero effect-size for far transfer effects (Melby-Lervåg et al., 2016), and recent studies using Bayesian statistics, also claim for evidence in favor of the null hypotheses (Clark, Lawlor-Savage, & Goghari, 2017).

Despite the strong negative evidence regarding far-transfer effects for working memory training, the majority of studies have shown that training in a working memory demanding task generalizes to other working memory tasks with a similar structure. Current meta-analyses suggested that the near-transfer effect is reliable and replicable (Melby-Lervåg & Hulme, 2013; Melby-Lervåg et al., 2016; Soveri, Antfolk, Karlsson, Salo, & Laine, 2017), with some showing near transfer effects for working memory training holding for as much as a few months post-training (Shahar & Meiran, 2015). Although the tiny (or even absent) far transfer effects might be discouraging at first sight, we believe that the reliable near transfer effects hold promise, and understanding their nature might enhance them and even extend transfer breadth. For example, pertinent questions are how much training does one need to generate near transfer effects? And what mechanisms might allow participants to show improved performance in a novel, never-before performed task? This study tries to shed light on these questions.

For this aim, we trained participants on a working memory demanding task for 12 weeks. In the training task, participants were asked to randomly switch between two choice-Reaction Time (RT) tasks, classifying either the spatial-location or the content of a target, using one of two manual responses. Importantly, on each training session, we frequently changed the task-set (target stimuli, response keys and mapping thereof), requiring participants to adapt and perform a new task-set after only a few trials. To ensure high working memory load as well as wider coverage of the various WM functions, the current task-switching task was combined with an N-back procedure, asking participants to also act according to information that was presented N trials beforehand (N was adapted according to the participants' performance; Jaeggi et al., 2008).

The notion behind the combination of rapid changes in task material (stimuli and response keys), N-back and task-switching was to increase demands for the maintenance and updating of action rules in working memory. The task-switching component was introduced to obligate participants to form task-sets which are hierarchical mental constructs that hold together information related to a specific task (e.g., task-cues, targets, response-keys and their relationships). The fact that task information kept changing between blocks (new stimuli and response keys on every block), assured that the information that participants were using to construct the task-sets was working memory demanding, and not based on long-term representations. Finally, the N-back component assured that across trials, the information required for task performance was held in mind for a brief moment and then immediately replaced by new information. Therefore, this task was designed to require participants to be able to repeatedly form, maintain and update task-sets based on novel information held in working memory.

Our hope was that rapid changes in the task-set (stimuli and response-keys), combined with high control demands (due to the task-switching and N-back combination) would encourage participants to form abstract representations of the overall task structure. This representation could comprise of interlinked slots for holding stimuli and their associated responses. Theoretically, such abstract stimulus-response rules should allow participants to flexibly allocate novel targets and response-keys to a well-trained abstract stimulus-response association. This theoretical assumption is not restricted to procedural training tasks (it can explain for example stimulus-stimulus associations) and can explain, at least in part, how participants gain expertise in a specific task structure regardless of any specific task-set. The notion of forming abstract task representations can therefore account for the so –called near transfer effect, where participants show improvement on a novel task that is very similar to the training task (e.g., same instructions and trial sequence but with different response keys and stimuli). However, our theoretical assumption also suggests that the transfer effect should be very limited, and should affect only tasks that directly use the abstract representations that have been formed.

In designing our study, we considered Oberauer (2009) who suggested a distinction between procedural and declarative working memory processing. Declarative working memory was suggested to hold representations relevant to knowledge and facts (based on stimulus-stimulus associations), while procedural working memory was proposed to hold action rules (Oberauer et al., 2013; Souza et al., 2012; but see Barrouillet, Corbin, Dagry, & Camos, 2014 for different results). Training studies have mainly explored declarative working memory processing (e.g., N-back, Span tasks), where participants are asked to memorized the presentation order of a stimulus set (i.e., stimulus-location associations). These studies have shown near transfer effects to similar declarative working memory measures with novel stimuli (Melby-Lervåg & Hulme, 2013). Here, we did not aim to explore whether declarative and procedural representations are held in two different sub-components of the working memory system (an issue that is still debated). Instead, we aimed to explore whether working memory training is specific and general at the same time – in the sense that it allows participants to form highly abstract (general) representations, of the (very specific) procedure they are training at.

Our main assumption was that working memory training would lead participants to form highly abstract task representations (stimulus-response abstract slots) which would allow them to subsequently show better performance on unpracticed procedural choice-RT tasks. On the other hand, such abstract representations should not benefit participants when performing declarative working memory or reasoning tests that also demand working memory resources but do not tap the same representational structure. Some evidence for the fact that the procedural choice-RT tasks and declarative working memory tasks tap different processes/representations in this specific data set was reported in a previous study that performed correlational analysis using only pre-test data from the current study (Meiran, Pereg, Givon, Danieli, & Shahar, 2016). In that study, it was found that a factor explaining the shared variance among the procedural working memory tasks was weakly related ($r=.12$, ns) to a factor explaining the variance among the declarative working memory tasks.

In the current study, we compared the effect of procedural working memory training with an active control training (visual search task, where participants were asked to find a target in an array of distractors, tapping relatively early perceptual processes) and passive control (no training). Importantly, in both working memory and active-control training tasks participants were required on each trial to report a decision between two alternatives using a manual key response (i.e., 2-alternative forced choice). However, stimulus-response associations were unchanged across the entire training in the visual search group, while for the working memory training group, participants had to adapt to frequent changes in the task-set across blocks (i.e., new stimuli and response keys) and trials (task-switching). Therefore, we assumed that if indeed participants can form abstract stimulus-response associations, this might be more strongly encouraged in the working memory training task. To assess training benefits, we measured performance in three types of working memory demanding transfer tasks: (1) reasoning, (2) declarative working memory, and (3) procedural working memory, assuming that procedural working memory training would benefit participants only in the latter. To assess dosage effects we also administrated procedural working memory transfer tasks in three additional time points in the study (after 2, 5 and 9 training sessions). Finally, we applied a mechanistic-based modeling approach to explore the choice mechanism that might underlie the observed transfer effects.

2. Methods

2.1 Participants

175 participants were recruited from the pre-academic preparatory course for engineering at Ben-Gurion University of the Negev over the course of two academic years. Participants were randomly assigned into training (N=72, 12 females, mean-age=23.32), active control (N=71, 12 females, mean-age =23.63) or silent control (N=32, 5 females, mean-age =23.81). Three participants (1 training, 2 silent control) dropped-out from the study and were omitted from analysis. Participants were given a monetary reward for completing the study.

2.2 Procedure overview

We recruited participants using mass-distributed emails using Ben-Gurion University approved publicity channels. Figure 1 describes the timeline for the study. All participants attended pre and post-test measurements (including procedural and declarative working memory, and reasoning testing). Training sessions were performed twice a week (~30 min each). Due to technical issues, Year I cohort completed 14 sessions and Year II completed 12 sessions (equally for training and active control subjects). Training was performed in consecutive weeks with no break. Sessions 3, 6 and 10 started with a novel procedural working memory transfer task (Shape RT Task), each testing employing new stimuli and stimulus-response mapping. The whole training was completed within 9-10 weeks for Year I and 8-9 weeks for Year II cohort. The silent control group did not practice during the time interval between pre and post-test. All testing and training sessions took place in a designated classroom located at Ben-Gurion University campus. The class contained 14 testing positions, each comprising of a desk and a desktop computer.

--- Insert Figure 1 about here ---

2.3 Training Tasks

2.3.1 Procedural working memory training task. The task was similar to that used in Shahar and Meiran (2015) and required participants to switch between two 2-choice RT tasks: Object classification task (report the target identity), and Spatial classification task (report the target location). In each training session, participants performed nine blocks of the task-switching task, each time using a new set of stimuli and responses. The task-set (stimuli and responses) was randomly composed each time by the computer from a set of 12 object pairs, 6 location pairs and 7 possible response pairs (see SM for further specifics).

Each block started with an instruction screen presenting the new task-set mappings (Figure 2, panel A for an example of an instruction screen). After the instruction screen, participants were given a chance to practice each task separately, across 4 trials for each task. This was followed by a reminder of the task-set mappings (Figure 2, panel A) and a test subblock (10 trials) in which the two tasks were randomly switched. Each trial sequence included a fixation (1,000 ms), task cue (500 ms), a second fixation point (1,000 ms) and the target stimulus (presented until a response or until 6,000 ms had elapsed). A 400ms beep signaled errors.

To make the task even more demanding in terms of working memory load, participants were asked in each trial to react according to information presented N trials beforehand. If the participant performed the test block with zero to one errors (.9-1 accuracy rate) the N-level increased by one. If they had four or more errors, the N-level was reduced. Otherwise, the N-level remained unchanged. On each block, the computer randomized whether the participant should perform an N-back for the cues or for the targets (see Figure 2). For example, a participant had an N=1 referring to the cues in a given block. This meant that in a given trial the participant would need to identify which task to perform according to the cue that was presented in the previous trial, and perform that task on the currently presented target. If the participant had to perform an N=1 referring to the target, the participant would need to identify the task according to the cue presented in the current trial, but respond to the target that was presented in the previous trial.

--- Insert Figure 2 about here ---

2.3.2 Visual search training task. This task served as an *active-control task* adapted from Redick et al., (2013). On each trial, participants were asked to report whether a target letter ('F') was facing right or left, using a right or left keypress respectively. Level of difficulty (array size and distractors composition) was set according to the participants' performance in the previous block. Each trial began with a 500ms fixation, followed by a 500ms letters-array, and then a 2,500ms mask during which the response was made. The participants' performance determined the difficulty level in two aspects – array size and composition. Odd-numbered levels were composed of trials with homogenous distracters (E's or T's facing left or right) and even-numbered levels were composed of trials with heterogeneous distracters (E's and T's facing left or right). The size of the array was altered every other level, such that the size would be 4 (2X2) on level 1 (homogenous), and 4 on level 2 (heterogeneous), and then 16 (4X4) on level 3 (homogenous) and 16 on level 4 (heterogeneous) etc. The size increase was made by adding two additional rows and columns. The criterion for level increase was accuracy higher than 87.5%, and the criterion for level decrease was accuracy lower than 75%. Otherwise, difficulty level remained unchanged. Each training session consisted of 16 blocks, with 24 trials, each. Since some participants managed to reach the maximal level of difficulty (100 letters, heterogeneous) we further increased the difficulty by increasing the requested accuracy: the criterion for level increase was accuracy higher than 92%, and the criterion for level decrease was accuracy lower than 80% (whereas the array was kept at 100 letters, heterogeneous).

2.4 Transfer Measurements

2.4.1 Procedural working memory (near transfer)

A battery of 6-choice-RT tasks was used. In each task, participants were asked to follow a set of six stimulus-response rules, and identify a target using a manual response (Shahar, Teodorescu, Usher, Pereg, & Meiran, 2014) . Three tasks were used:

(1) Letter and digit classification tasks – These tasks included two Mapping conditions (arbitrary vs. non-arbitrary). In the arbitrary mapping condition, stimulus-response rules were novel and thus needed to be maintained in WM. The non-arbitrary mapping was based on familiar knowledge assumed to be well represented in long-term memory and thus considered to place little demand (or no demand) on WM. Non-arbitrary stimulus-response rules were letters arranged alphabetically and digits

arranged by their numeric size from left to right. We used two different sets of arbitrary rules- one for pretest and one for post-test, to ensure that individual differences in each measurement (pre-test, post-test) are not influenced by differences between the arbitrary rules**.**

(2) Shape classification task – This task included only arbitrary mapping which was performed during pre-test, post-test as well as three times during training (at the beginning of Sessions 3, 6, & 10). On each measurement, a novel set of stimuli appeared.

In all three tasks, participants responded using six horizontally adjacent keyboard keys on the lower row of the keyboard (i.e., x,c,v,b,n,m in a QWERTY keyboard). At the beginning of each block, participants were presented with an instruction screen specifying which key should be pressed in response to each target presentation (6 sec deadline). Following the instructions screen, a short practice phase was administrated, including one test trial for each stimulus-response rule (i.e., six trials). Following the practice phase, the instruction screen presenting the stimulus-response rules for the current condition reappeared, giving participants a second chance to memorize the current mapping. Afterward, participants completed a test phase of 100 trials. Each trial consisted of a fixation (250ms), target (until response or until 6s had elapsed) and a blank screen inter-stimulus interval (250ms). A 400ms beep signaled errors.

2.4.2 Declarative working memory (Intermediate Transfer Measurements)

Declarative working memory was estimated using two Automated Complex Span Tasks (Unsworth, Heitz, Schrock, & Engle, 2005). In these tasks participants are asked to memorize and retrieve a series of stimuli in the order they were presented. Between each stimulus presentation participants were asked to perform a secondary task aimed to prevent phonological rehearsal. Two tasks were used: (1) Operation-Span - memorizing letters while solving simple math equations, and (2) Symmetry span – memorizing spatial locations while performing a symmetry judgement. The same tests were given at post-test. The E-Prime ™ versions for the standard tasks were downloaded from http://englelab.gatech.edu/tasks.html and the instructions in them were translated to Hebrew. See also Supplementary Material.

2.4.3 Reasoning (Far Transfer Measurements)

Two paper and pencil tests were used to estimate reasoning abilities:

(1) ETS-Locations: This test requires to find and then apply rules. Each item includes Xs, dashed lines and spaces appearing in spatial locations according to a certin rule. Participants are required to find the rule and indicate the next location of the X. Forms A and B were used for pre-test and post-test, respectively. Time was limited to six minutes. Each form included 14 test-items. This is a subtest from the ETS-Kit Induction factor (Ekstrom, French, Harman, & Derman, 1976). See also Supplementary Materials.

(2) Following instructions test (Verbal): On each item participants are requested to follow highly complex instructions (e.g., "In the following digit sequence, count how many times does the digit 7 appear after an even number"). We divided the test into two equal parts (odd and even items, 20 items each), to be used for pre and posttest measurements. We used a seven minutes time limit. This is a subtest in the Intellectual Differential Aptitude Test battery (Fischman, 1982).

3. Results

We now review our results starting with training effects (how much participants improved in the training task). We next review the transfer effects (pre vs. post) in each of our three transfer domains: procedural working memory, declarative working memory and reasoning. We then report the dosage effect of training on transfer to a procedural working memory task. Finally, evidence accumulation modeling was performed to explore the choice mechanisms underlying the observed transfer effects. We employed Bayesian statistical inference in all analyses (using JASP; Love et al., 2015), adopting equal priors for H0 and H1.

3.1 Training effect.

Both the training and the active-control groups showed substantial improvement in the training level (N value / distractor set-size) as reflected in a large and statistically significant effect of Session on mean training level (η_p^2=.56, BF_{10}=$2e^{263}$).

3.2 Transfer effects (pre vs. post-test)

3.2.1 Procedural working memory (near transfer).

Here we analyzed the performance of the training, active-control and passive-control groups in the letter and digit-classification transfer tasks. These were 6-choice RT tasks, performed under high and low working-memory load (manipulated by means of the mapping arbitrariness, i.e., task novelty vs. familiarity).

For reaction-times analyses, error and post-error trials, ten first trials in each condition and the first trial after each recess were discarded. Reaction-times below 200ms or above 3.5 SDs from the participant’s mean in the respective condition were considered as outliers and thus omitted

(Schmiedek et al., 2007; Shahar et al., 2014). For the mean-RT, we found strong evidence for lack of transfer effects (see Figure 3). This was reflected in a small Bayes Factor for the Group (3; training, active-control, passive-control) x Time (2; pre vs. post) interaction (BF_{10}=.03, η^2_p=.01), indicating that the null hypothesis was 33 times more probable than H1. A similar result was found concerning the triple Group (3; training, active-control, passive-control) x Time (2; pre vs. post) x Load (2; arbitrary vs. non-arbitrary) interaction (BF_{10}=.05, η^2_p=.01).

In accuracy rates, we found strong evidence in favor of a Group (3; training, active-control, passive-control) x Time (2; pre vs. post) interaction (BF_{10}=2.7e6, η^2_p=.14), showing higher accuracy rates in post-test for the training group compared to the two control groups (see Figure 3). There was clear evidence for a null three-way Group (3; training, active-control, passive-control) x Time (2; pre vs. post) x Load (2; arbitrary vs. non-arbitrary) interaction (BF_{10}=.05, η^2_p=.02), indicating that the null hypothesis was 20 time more likely than H1, given the results. This finding suggests that the Group x Time interaction did not change with working-memory load.

--- Insert Figure 3 here ---

3.2.2 Declarative working memory (far transfer). We calculated a Bayesian ANOVA with Task (2; Operation-span vs. Symmetry-span), Time (2; pre vs post) and Group (3; training, active-control, passive-control) as predictors and Complex Span score as dependent. Group x Time interaction showed evidence in favor of the null hypothesis (BF_{10}=.04, η^2_p<.01) suggesting no training benefits for declarative working memory (see Figure 4).

3.2.3 Reasoning (far transfer). We calculated a Bayesian ANOVA with Task (2; Following instructions test vs. ETS-locations), Time (2; pre vs post) and Group (3; training, active-control, and passive-control) as predictors and reasoning score as dependent. Group x Time interaction showed evidence in favor of the null hypothesis (BF_{10}=.04, η^2_p<.01) suggesting no training benefits for reasoning (see Figure 4).

--- Insert Figure 4 here ---

3.3 Dosage effects. Here we analyzed the performance of training vs. active-control in the shape-classification transfer tasks. These 6-choice RT tasks involved novel (arbitrary) stimulus-response mappings and was administrated at five time points – pre-test, start of session 3, start of session 6, start of session 10 and post-test. Thus, two independent variables were included in the analysis: Time (5 time points) and Group (2; training vs. active control). In accuracy-rates, we found strong evidence in favor of a paired Group x Time interaction (BF_{10}=5.06e5, η^2_p=.07). The numerical effect shows higher accuracy for the training group emerging as soon as after two training sessions (see Figure 5). To explore the magnitude of the training effect after two training sessions, we re-examined the Group x Time interaction only on the two first time-points (2; pre-test vs. start of Session 3). We found strong evidence in favor of the paired Group x Time interaction (BF_{10}=71.75, η^2_p=.09). Note, that while it may look like the Group x Time effect is the result of reduced performance for the active control across time, one cannot make such a clear interpretation. The reason is that a different set of stimuli were used for each measurement, which can result in different overall task difficulty for each time measurement. In reaction-times, we found strong evidence against a Group (2; training vs. control) x Time (2; pre vs. post) interaction (BF_{10}=.02, η^2_p<.01).

--- Insert Figure 5 here ---

3.4 Evidence Accumulation Modeling.

To explore for the mechanism underlying the effect of training on accuracy and RTs, we fitted the two-stage-decision model to choice-RT data (Shahar, Teodorescu, Anholt, Karmon-Presser, & Meiran, 2017; Shahar et al., 2016, 2014). The two-stage-decision model is an evidence accumulation model describing the decision as involving two processing stages. The first stage involves identification/classification of the target (e.g., knowing that the identity of a target such as a number or letter). This stage is modeled using a max-minus-next mechanism (Krajbich & Rangel, 2011; Teodorescu & Usher, 2013). In max-minus next (Teodorescu & Usher, 2013), N accumulators race independently until the gap between the highest and the second highest accumulator reaches a threshold. The two most relevant parameters of this model include *drift rate* – which is the rate of evidence entering the accumulation mechanism (assuming fixed level of noise), and *response threshold* pertaining to the minimum difference between the two leading accumulators that will stop the accumulation process.

The second stage involves the retrieval of the action rule (i.e., knowing what response is associated with the target that was identified in the previous step). The second decision stage is extrinsic to the evidence accumulation mechanism and involves action-rule retrieval (Shahar et al., 2014). We hypothesized that given the very short duration, a reasonable approximation is that this process is characterized by a constant rate, i.e., a constant hazard function. Since a constant hazard function results in an exponential RT probability density function, the two-stage model adds an exponential non-decision time (*lambda*) to the decision RT generated by the max-minus next process (Shahar et al., 2014).

3.4.1 Models. We constructed two classes of models, the first class assuming that the training-related benefits involve the perceptual classification, described by the evidence accumulation process (Models 1-3). The second class of models assumed that training affected rule-retrieval rate, either with or without also affecting the evidence accumulation process (Model 4-6). Finally, we assembled a null model where we assumed no group differences (for comparison). Specifically, each model assumes that better performance for the training group compared with active control at post-test is due to:

- Model 1 – more efficient perceptual classification of the target as a result of higher quality of perceptual evidence going into the accumulation mechanism for the training group (higher drift rate coupled with lower response threshold for the training group).
- Model 2 – less fluctuations in motivation/early-attention across trials leading to more efficient perceptual classification of the target (higher drift-rate and lower drift variability parameters for training group)
- Model 3 –a more optimal response policy following training. It is argued that after some time without a decision, the agent can benefit from lowering the response threshold (Hawkins, Forstmann, Wagenmakers, Ratcliff, & Brown, 2015). That is, collapsing threshold was claimed to describe a more optimal decision policy. Therefore, in this model we assume that higher productivity for trained participants is the result of a quicker-to-collapse threshold.
- Model 4 – quicker retrieval of the correct action-rule (lower *lambda* parameter for the training group compared with active-control).

- Model 5 – a mixture model which assumes that both perceptual classification and retrieval of the action rule is improved following training (higher drift-rate, lower lambda for training group)
- Model 6 – a mixture model that assumes more careful response policy for perceptual classification and quicker retrieval of the action rule following training (higher threshold and lower *lambda* for training group)
- Model 7 – Null model assuming no group differences to allow a basis for comparison.

3.4.2 Fitting procedure. The models were fitted to the working memory training and active-control groups data at post-test (only high load condition, across the three choice-RT tasks). For model fitting, we pooled the data across participants and tasks in each group. To reduce variance between participants and sessions, we first subtracted from each RT the median score of each participant's respective condition. We then added to each RT the median score of the respective group (active control vs. training) calculated across participants and measurements. We next calculated the .1 .3 .5 .7 .9 quantiles for each group across participants and conditions, once for correct and once for error choices. We calculated the quantile probability function (QPF) by multiplying the RT quantiles with the correct / error rates respectively. This resulted in a probability score representing the chance of each response (error / correct) to fall within each quantile. Models were then fitted with the quantile probability function. Models were also compared with respect to their Bayesian Information Criterion (Raftery, 1995) with smaller BIC indicating better fit.

3.4.3 Modeling results. Results showed that the model with both improved drift-rate and quicker retrieval time for the training group (Model 5) provides the best BIC scores (see Table 1, for parameters recovery see Appendix A). In fact, this model outperformed the next-best fitting model by 47.8 BIC units, meaning that the likelihood of the chosen model, given the results is more than 2e10 times greater than that of the next-best model. Therefore, according to the two-stage model results, the improved performance after training can be attributed to more efficient classification of the target and quicker retrieval of the action rules. We should note that the pooling method is not optimal and was selected due to the fact that this model cannot be fitted to participants who did not commit any errors. However, we are not aware of any specific problems or biases this approach might have, and believe the selected sub-set of data best represented the effect reported in the accuracy and RT analysis in the previous section.

Table 1. Modeling results

	Model	BIC	Free parameters
1	Drift-rate & response threshold	156427.49	9
2	Drift-rate & drift variability	156525.93	9
3	Collapsed threshold	156443.24	13
4	Retrieval time	156419.29	8
5	**Retrieval time & drift-rate**	**156371.49**	**9**
6	Retrieval time & response threshold	156433.37	9
7	Null model	156564.86	7

4. Discussion

In the current study, we examined whether working memory training can improve performance in untrained tasks with a similar structure (near-transfer effect). We were interested in two questions: What dosage might be required to induce training benefits, and what choice-mechanisms might underlie such an effect? For this aim, we randomly assigned participants to one of three groups: procedural working memory training, visual search training (active control) and no-contact (passive control). Both working memory and visual search training groups performed 2-alternative forced choices, however only the working memory training demanded that participants adapt to a new set of targets and response keys and was performed under high cognitive control demands (switching tasks and reacting according to the target in the N-back trial).

We tested for near transfer effects using procedural working-memory tasks, which were a set of 6-choice RT tasks with novel stimuli. When compared with control participants, trained participants exhibited higher accuracy-rate in the 6-choice RT task. This effect was noticeable as quickly as after two training sessions, suggesting that very little practice is needed in order to produce near-transfer effects. While the working memory training group did not exceed the control groups in terms of reaction-times, there was also no evidence for a speed-accuracy trade-off. That is, while the working memory training group had better accuracy rates, this was not the result of a more cautious response policy. Our measurement battery also included two types of 6-choice RT tasks performed under either high vs. low working memory demands (arbitrary vs. non-arbitrary mapping). Our results suggest that the working memory group had better accuracy rates in both high and low working memory 6-choice RT transfer tasks, suggesting that the training had an effect on some general property of the choice-RT related processes.

What choice mechanism might underlie this near transfer effect? To try and answer this question we compared mechanisms for two relevant processes: better classification of the target, and a quicker retrieval of action rules. We found that the best model was the one explaining training benefits in terms of higher perceptual sensitivity (higher drift-rates) and quicker retrieval of action rules (*lambda* parameter). While we did assume an effect of working memory training on action rule retrieval latencies, we did not assume a higher drift rate following training. According to our own two-stage model evidence accumulation mechanism is mostly influenced by the perceptual identification process (Shahar et al., 2014). Therefore, this result (given that our two-stage model is not completely wrong) suggests that working memory training not only improved the retrieval of the responses, but also improved the ability to reach a fast and accurate classification of the target. Holding in mind that the training task used a two-choice paradigm, while the transfer tasks were designed with six-choices, we suggest that the observed transfer effect might reflect a highly abstract representation of stimulus-response rules in general. We hypothesize that this effect can be explained under a limited attention resources assumption: novel arbitrary choice-RT requires attentional resources. If the process of rule retrieval improved for trained participants, it might be less demanding and therefore it frees up resources, which in turn allows a more efficient early perceptual classification of the target. This suggests that training is driving participants to optimize their resource allocation and usage in a given context, without changing the overall amount of available resources (Oberauer & Kliegl, 2004). However, this is of course only speculative and should be tested in a designated design.

Next, we tested for an intermediate transfer effect using a set of declarative working memory tasks. While these were working memory demanding tasks, their task structure was very different than that of the training task. What should we expect in terms of transfer to an untrained working memory task with a different structure? If training did not improve participants overall available resources for working memory processing, we should see very little transfer to the declarative working memory measurements. Indeed, we did not observe any such transfer. Accordingly, Reasoning scores, a measure of far transfer, were also unaffected by working memory training. While reasoning tasks are known to be working memory demanding (Süß et al., 2002), they shared no common features with the training task. Our result considering lack of far transfer effect are in-line with previous studies (Lindenberger et al., 2017; Melby-Lervåg et al., 2016). Working memory training abilities do not seem to improve outside the context of the specific paradigm on which it was trained. Our Bayesian analysis allow us to also quantify the amount of evidence in favour of the null hypothesis, and indeed, the intermediate and far transfer results enabled us to accept the null hypothesis. The fact that other working memory demanding transfer tasks (declarative working memory, reasoning) did not show transfer effects, suggests that this is somewhat of a "specific generalized" effect – the training effect is generalized to other untrained tasks with different stimuli, responses and set-size, but only to the extent that the transfer task uses a similar overall structure.

Following the training literature, it seems unreasonable to assume that training affects the amount of resources available for task performance. Rather, it is much more plausible to assume that transfer effects represent a more efficient use of the same resources (Oberauer & Kliegl, 2004). Our main assumption is that the highly demanding nature of the training task (holding N items in memory and switching between tasks), forced participants to form highly abstract representations of stimulus-response task structure, allowing them to flexibility use this resource when performing other untrained choice-reaction tasks, even when other stimuli and responses are used. Modelling result seem to go in line with this suggestion (showing higher *lambda*).

Before concluding, we would like to mention that the only published study that used a similar procedural working memory task was from our group (Shahar & Meiran, 2015). In this study, wefound transfer to choice-RT tasks in RT measurements, rather than accuracy-rates. While we do not have any data that could be used to bridge the difference in results, we can point to a few differences in the design that might have caused this discrepancy. First, Shahar &

Meiran (2015) had a much smaller sample and used only a passive control group. This might have caused us to overestimate the effect of training on RTs. In addition, Shahar & Meiran (2015) did not use exactly the same task, but rather a different version where the stimuli and response changed in a lower rate across the session, which might have reduced the effect, as it necessitates a less frequent use of this theoretical abstract task structure.

To sum up, we report improved performance in working memory demanding choice-RT transfer tasks following working memory training. The transfer effect was limited to choice-RT tasks, and seem to be specific to working-memory training but not visual attention training, despite both training tasks were essentially choice-RT tasks. Dosage analysis suggests that this effect appears after as little as two training sessions. The model fitting results suggest higher perceptual sensitivity for the target and quicker retrieval of action rules from working memory to underlie the observed transfer effect.

5. Limitations

We did not include near transfer measures that would have allowed us to explore for a near transfer effect in the active-control group. For example, we could have used a novel, untrained visual search task. In hindsight, this might have allowed us to see a context related near-transfer effect, for both groups. Therefore, we cannot conclude that these results are specific to working memory, leaving this question to future studies. Also, the lack of follow-up measurement prevents us from concluding to what extent this effect persists after training has stopped. Another issues is the lack that our dosage effect was assessed only by comparing training to active-control with no passive control (which was due to technical difficulties in the study administration). Finally, additional limitation is that the current working memory task (combining both task switching and N-back paradigms) was not used outside our group (Shahar & Meiran 2015, see Discussion), making it difficult to compare our results with previous ones.

Acknowledgements

We would like to thank Ella Givon and Gal Danieli for their help with managing this large-scale study. We also wish to thank Daniel Aranovich, Gal Berger, Dan Halunga, Ayelet Itzhak, Nadav Kozlovsky, Inbal Michel, Elad Naor, Liad Olansky, Hovav Paller, Elisha Puderbeutel and Adva Weinstein for their invaluable help with data collection.

References

Andreou, P., Neale, B. M., Chen, W., Christiansen, H., Gabriels, I., Heise, A., … Kuntsi, J. (2007). Reaction time performance in ADHD: improvement under fast-incentive condition and familial effects. *Psychological Medicine*, *37*, 1703–1715. doi:10.1017/S0033291707000815

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience*, *4*, 829. doi:10.1038/nrn1201

Barrouillet, P., Corbin, L., Dagry, I., & Camos, V. (2014). An empirical test of the independence between declarative and procedural working memory in Oberauer's (2009) theory. *Psychonomic Bulletin & Review*, 1–6. doi:10.3758/s13423-014-0787-y

Carruthers, P. (2013). Evolution of working memory. *Proceedings of the National Academy of Sciences*, *110*, 10371–10378. doi:10.1073/pnas.1301195110

Clark, C. M., Lawlor-Savage, L., & Goghari, V. M. (2017). Working memory training in healthy young adults: Support for the null from a randomized comparison to active and passive control groups. *PLOS ONE*, *12*, e0177707. doi:10.1371/journal.pone.0177707

Constantinidis, C., & Klingberg, T. (2016). The neuroscience of working memory capacity and training. *Nature Reviews Neuroscience*, *17*, 438–449. doi:10.1038/nrn.2016.43

Ekstrom, R. B., French, J. W., Harman, H., & Derman, D. (1976). Kit of factor-referenced cognitive tests (rev. ed.). *Princeton, NJ: Educational Testing Service*.

Fischman, E. (1982). IDAT: Intellectual differential aptitude test battery. Hamerkaz Lekhinukh Technologi, Holon, Israel. Hebrew–English abstract.

Hawkins, G. E., Forstmann, B. U., Wagenmakers, E.-J., Ratcliff, R., & Brown, S. D. (2015). Revisiting the evidence for collapsing boundaries and urgency signals in perceptual decision-making. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *35*, 2476–2484. doi:10.1523/JNEUROSCI.2410-14.2015

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

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. *Psychonomic Bulletin & Review*, *9*, 637–671. doi:10.3758/BF03196323

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

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences*, *108*, 13852–13857. doi:10.1073/pnas.1101328108

Lindenberger, U., Wenger, E., & Lövdén, M. (2017). Towards a stronger science of human plasticity. *Nature Reviews Neuroscience*, *18*, 261–262. doi:10.1038/nrn.2017.44

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A. J., … Wagenmakers, E. J. (2015). JASP (Version 0.6.6) (Version Version 0.7.5).

Meiran, N., Pereg, M., Givon, E., Danieli, G., & Shahar, N. (2016). The role of working memory in rapid instructed task learning and intention-based reflexivity: An individual differences examination. *Neuropsychologia*, *90*, 180–189. doi:10.1016/j.neuropsychologia.2016.06.037

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*, 270–291. doi:10.1037/a0028228

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." *Perspectives on Psychological Science*. doi:10.1177/1745691616635612

Oberauer, K. (2009). Chapter 2 Design for a working memory. In Brian H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 51, pp. 45–100). Academic Press.

Oberauer, K., & Kliegl, R. (2004). Simultaneous Cognitive Operations in Working Memory After Dual-Task Practice. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 689–707. doi:10.1037/0096-1523.30.4.689

Oberauer, K., Souza, A. S., Druey, M. D., & Gade, M. (2013). Analogous mechanisms of selection and updating in declarative and procedural working memory: Experiments and a computational model. *Cognitive Psychology*, *66*, 157–211. doi:10.1016/j.cogpsych.2012.11.001

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–163. doi:10.2307/271063

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. *Journal of Experimental Psychology: General*, *142*, 359–379. doi:10.1037/a0029082

Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H.-M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General*, *136*, 414–429. doi:10.1037/0096-3445.136.3.414

Shahar, N. (2018). Formation of abstract task representations: Exploring dosage and mechanisms of working memory training effects. Retrieved from osf.io/f5j94.

Shahar, N., & Meiran, N. (2015). Learning to Control Actions: Transfer Effects following a Procedural Cognitive Control Computerized Training. *PLoS ONE*, *10*. doi:10.1371/journal.pone.0119992

Shahar, N., Teodorescu, A. R., Anholt, G. E., Karmon-Presser, A., & Meiran, N. (2017). Examining procedural working memory processing in obsessive-compulsive disorder. *Psychiatry Research*, *253*, 197–204. doi:10.1016/j.psychres.2017.03.048

Shahar, N., Teodorescu, A. R., Karmon-Presser, A., Anholt, G. E., & Meiran, N. (2016). Memory for action rules and reaction-time variability in Attention-Deficit/Hyperactivity disorder. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *0*. doi:10.1016/j.bpsc.2016.01.003

Shahar, N., Teodorescu, A. R., Usher, M., Pereg, M., & Meiran, N. (2014). Selective influence of working memory load on exceptionally slow reaction times. *Journal of Experimental Psychology: General*, *143*, 1837–1860. doi:10.1037/a0037190

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, *138*, 628–654. doi:10.1037/a0027473

Souza, A. da S., Oberauer, K., Gade, M., & Druey, M. D. (2012). Processing of representations in declarative and procedural working memory. *Quarterly Journal of Experimental Psychology (2006)*, *65*, 1006–1033. doi:10.1080/17470218.2011.640403

Soveri, A., Antfolk, J., Karlsson, L., Salo, B., & Laine, M. (2017). Working memory training revisited: A multi-level meta-analysis of n-back training studies. *Psychonomic Bulletin & Review*, *24*, 1077–1096. doi:10.3758/s13423-016-1217-0

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*, 261–288. doi:10.1016/S0160-2896(01)00100-3

Teodorescu, A. R., & Usher, M. (2013). Disentangling decision models: From independence to competition. *Psychological Review*, *120*, 1–38. doi:10.1037/a0030776

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498–505. doi:10.3758/BF03192720

Wilhelm, O., & Oberauer, K. (2006). Why are reasoning ability and working memory capacity related to mental speed? An investigation of stimulus–response compatibility in choice reaction time tasks. *European Journal of Cognitive Psychology*, *18*, 18–50. doi:10.1080/09541440500215921

Figures

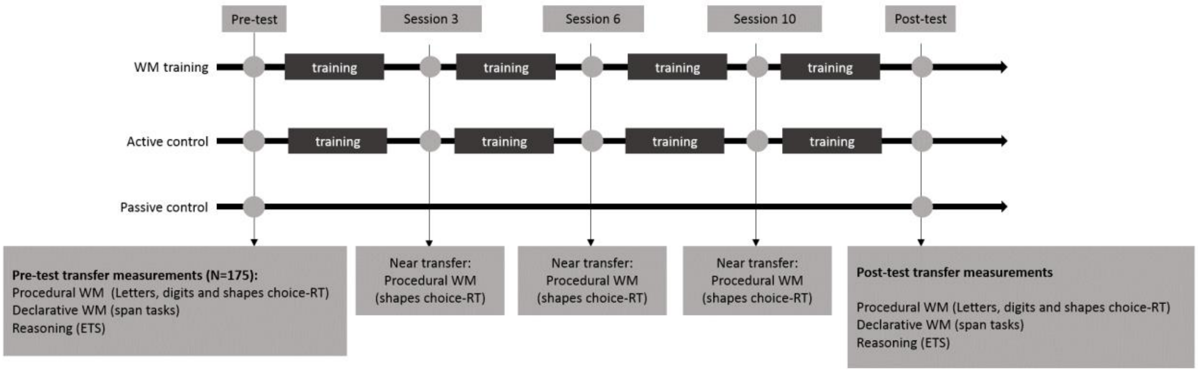

Figure 1. Timeline for the training study.

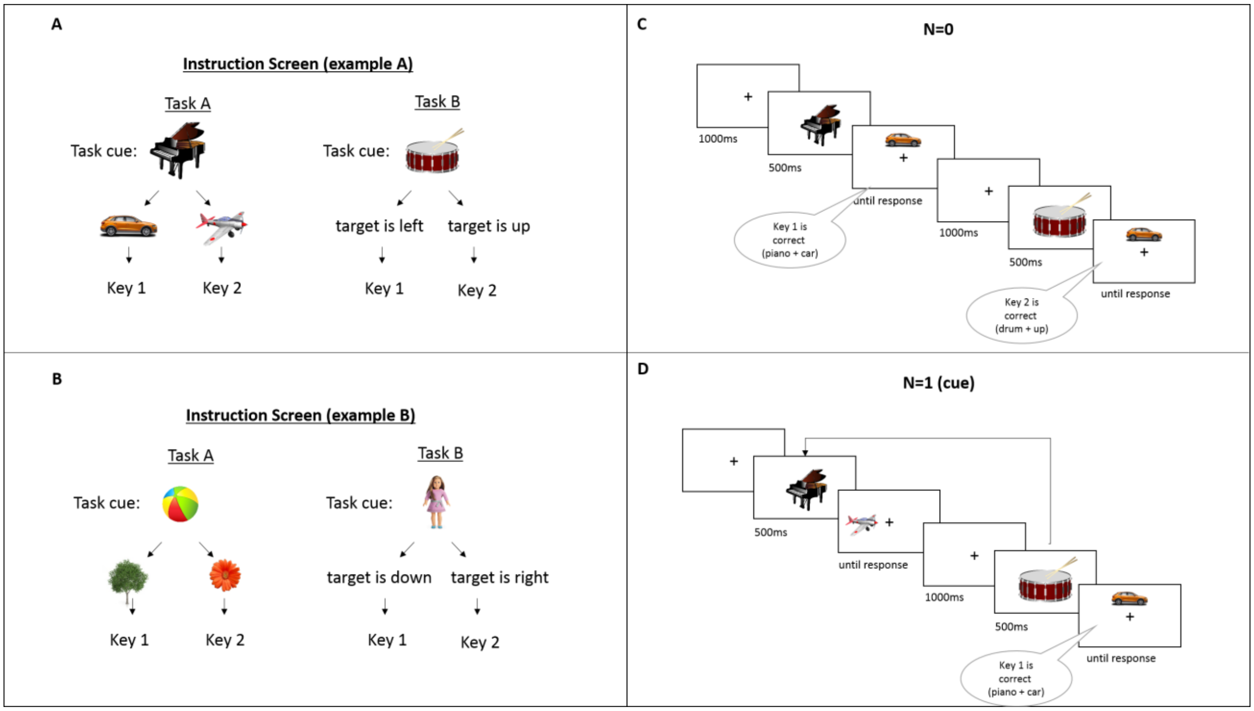

Figure 2. Illustrations for the working memory training task, where participants performed a task-switching task. Each block participants were asked to memorize a new task-set for these two tasks. Panel A and B show an example for two random task-sets, showing task-cues for the Object and Spatial tasks (piano/drum, panel A and ball/doll, panel B) related target-response rules. Panel C illustrates a task sequence according to the instructions shown in Panel A: piano=object task, then press the key related to 'car'. Drum=spatial task, then press the key related to 'up'. Panel D illustrates the same task-set, with N=1. Here participants needed to perform a task in the current trial, according to the task-cue presented N trials beforehand.

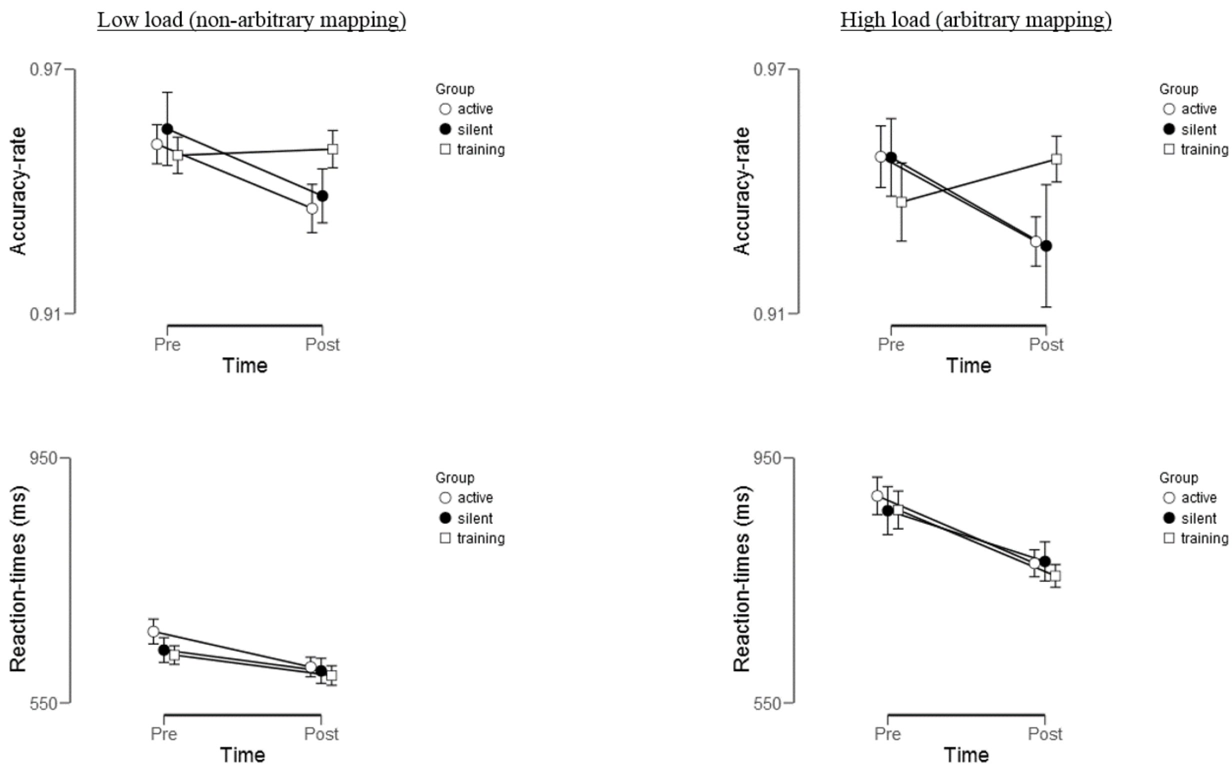

Figure 3. Accuracy rates (upper panels) and mean RT (lower panels) in the choice-RT transfer tasks as a function of Group, Time and mapping arbitrariness. Results are collapsed across the letter and digit-classification tasks. Error bars represent 95% credible intervals.

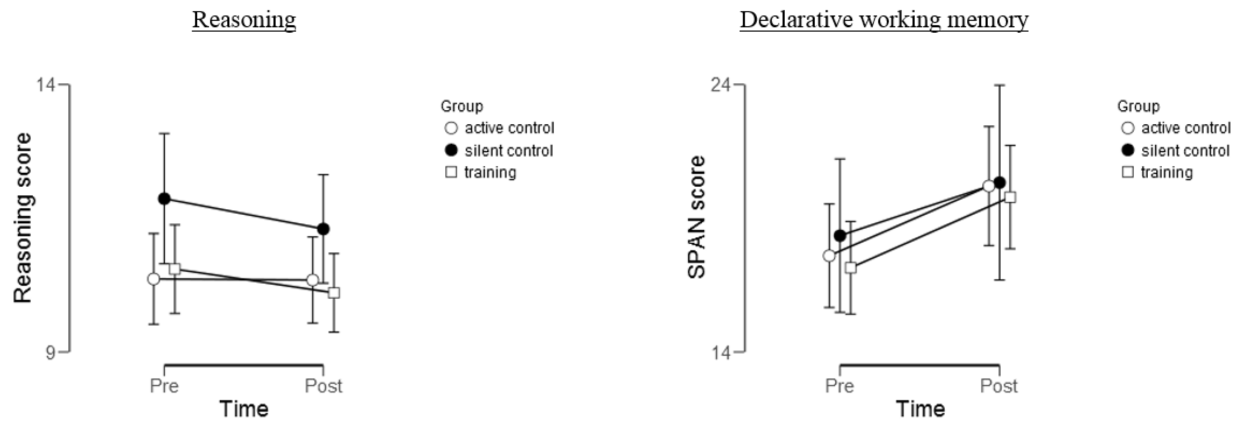

Figure 4. Transfer effect for reasoning (left panel) and declarative working memory (span tasks, right panel) for the training and control groups. Error bars represent 95% credible intervals.

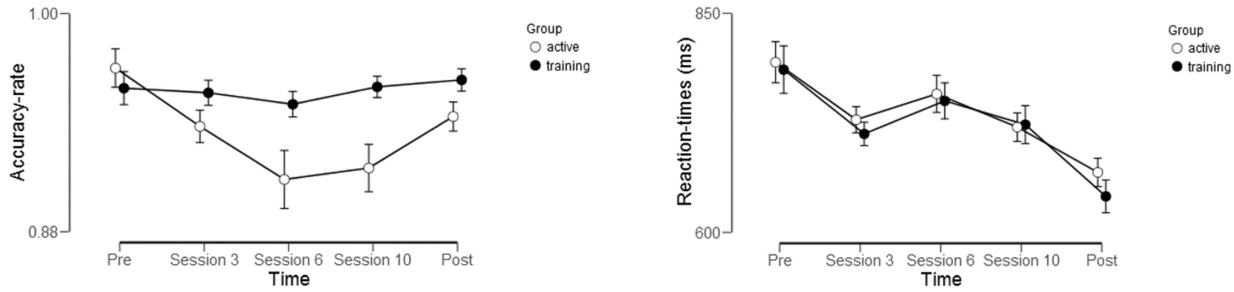

Figure 5. Accuracy rates (left panel) and mean RT (right panel) in the 6-choice-RT transfer tasks as a function of Group and Time.