International Psychogeriatrics: page 1 of 15 © International Psychogeriatric Association 2015
doi:10.1017/S1041610215001519

Training of resistance to proactive interference and working memory in older adults: a randomized double-blind study

Sandra V. Loosli,[1,2,3] Rosalux Falquez,[1,4] Josef M. Unterrainer,[5] Cornelius Weiller,[1,2,6] Benjamin Rahm[5] and Christoph P. Kaller[1,2,6]

[1] *Department of Neurology, University Medical Center Freiburg, Freiburg, Germany*
[2] *Freiburg Brain Imaging Center, University of Freiburg, Freiburg, Germany*
[3] *Biological and Personality Psychology, Department of Psychology, University of Freiburg, Freiburg, Germany*
[4] *Department of Clinical Psychology and Psychotherapy, University of Heidelberg, Heidelberg, Germany*
[5] *Medical Psychology and Medical Sociology, University Medical Center Mainz, Mainz, Germany*
[6] *BrainLinks-BrainTools Cluster of Excellence, University of Freiburg, Freiburg, Germany*

ABSTRACT

Background: Working memory (WM) performance is often decreased in older adults. Despite the growing popularity of WM trainings, underlying mechanisms are still poorly understood. Resistance to proactive interference (PI) constitutes a candidate process that contributes to WM performance and might influence training or transfer effects. Here, we investigated whether PI resistance can be enhanced in older adults using a WM training with specifically increased PI-demands. Further, we investigated whether potential effects of such a training were stable and entailed any transfer on non-trained tasks.

Method: Healthy old adults ($N = 25$, 68.8 ± 5.5 years) trained with a recent-probes and an *n*-back task daily for two weeks. Two different training regimens (high vs. low PI-amount in the tasks) were applied as between-participants manipulation, to which participants were randomly assigned. Near transfer tasks included interference tasks; far transfer tasks assessed fluid intelligence (gF) or speed. Immediate transfer was assessed directly after training; a follow-up measurement was conducted after two months.

Results: Both groups similarly improved in PI resistance in both training tasks. Thus, PI susceptibility was generally reduced in the two training groups and there was no difference between WM training with high versus low PI demands. Further, there was no differential near or far transfer on non-trained tasks, neither immediately after the training nor in the follow-up.

Conclusion: PI-demands in WM training tasks do not seem critical for enhancing WM performance or PI resistance in older adults. Instead, improved resistance to PI appears to be an unspecific side-effect of a WM training.

Key words: working memory, proactive interference, inhibition, training, intervention, transfer, aging, fluid intelligence

Introduction

Working memory (WM) as the ability to simultaneously store and manipulate information in mind (Baddeley, 1997) is a key foundation of human cognition (Daneman and Carpenter, 1980; Kyllonen and Christal, 1990). Trainings of WM have become increasingly popular during the last decade with older adults being one of the main target groups for cognitive interventions. Several studies in older adults reported transfer after WM training not only on untrained WM measures, but also on other cognitive functions such as inhibition (e.g. Borella *et al.*, 2010), gF (e.g. Carretti *et al.*, 2013a; Zinke *et al.*, 2014), or language comprehension (Carretti *et al.*, 2013b). Although these results seem promising, the reported transfer effects are inconsistent (see e.g. Shipstead *et al.*, 2010; Melby-Lervåg and Hulme, 2013; Redick *et al.*, 2013). Even more importantly, the mechanisms that underlie successful training or transfer still remain unclear (e.g. Jolles and Crone, 2012; Melby-Lervåg and Hulme, 2013). Recent studies showed that properties of the training regimen might influence training success (e.g. performance-adaptive vs. non-adaptive training tasks, see e.g. Klingberg *et al.*, 2005, or the duration

Correspondence should be addressed to: Lic. phil. Sandra Loosli, Department of Neurology, University Medical Center, Breisacher Str. 64, 79106 Freiburg, Germany. Phone: + 49-761-270-54794; Fax: + 49-761-270-54819. Email: sandra.loosli@uniklinik-freiburg.de. Received 23 Feb 2015; revision requested 17 Jul 2015; revised version received 11 Aug 2015; accepted 19 Aug 2015.

of the intervention, see Jaeggi *et al.*, 2008). Also characteristics of the participants might influence improvements in training or transfer tasks, such as age (Wass *et al.*, 2012; Borella *et al.*, 2014), baseline performance and gain in the training tasks (Zinke *et al.*, 2014) or individual differences in motivational factors or personality traits (see Studer-Luethi *et al.*, 2012; Jaeggi *et al.*, 2014). But as of yet, only little is known about cognitive processes underlying successful WM training and transfer.

Here, we aimed to investigate the relationship between WM training and the ability to resist PI – a process that has repeatedly been shown to account for a substantial portion of inter-individual variation in WM capacity (Mecklinger *et al.*, 2003; Jonides and Nee, 2006). PI refers to a reduced recall accuracy and prolonged reaction time (RT) due to previously relevant but now irrelevant WM contents (Postle *et al.*, 2004). PI can occur in all sorts of WM tasks, i.e. not only in simple and complex span tasks (May *et al.*, 1999; Rowe *et al.*, 2008) that are based on free recall but also in *n*-back (e.g. Burgess *et al.*, 2011) or recent-probes tasks (Jonides and Nee, 2006) that are based on recognition. Building on the significance of PI resistance for the reliability of WM operations, Persson and Reuter-Lorenz (2008) investigated if resistance to PI can be trained and to which extent a potential improvement would transfer. To this end, two groups of young adults were compared that followed almost identical training regimens with the only difference that the WM tasks of the experimental group included trials with PI, while the tasks of the active control group contained less or no such trials. In the study of Persson and Reuter-Lorenz (2008), there were actually three different training groups: An experimental group, which trained on the WM tasks with high interference, a control group 1, which trained on the same WM tasks, however with no or low interference demands, and a control group 2, which trained on similar tasks with no demands on interference resolution and only minimal demands on WM. As there was no difference between both control groups before or after training, they were put together for the analyses (Persson and Reuter-Lorenz, 2008). Results revealed considerable improvements and transfer following training with increased demands on PI resistance but had to be retracted due to experimenter error (Persson and Reuter-Lorenz, 2011). In their comment accompanying the retraction, Persson and Reuter-Lorenz (2011) noted that after their research was publicized, they detected a mistake in the programming of the training tasks. Further, they stated that in a separate replication that used randomized, novel sequences, they could not replicate the findings of the first report. In another study with young participants that manipulated interference control in an *n*-back task, transfer to non-trained PI-tasks was not investigated, and only partial transfer was reported to a cued version of a flanker task, while transfer to a complex span task and to matrix reasoning did not differ from a passive control group (Oelhafen *et al.*, 2013). Further, Bomyea and Amir (2011) conducted an intervention that consisted only of a single training session with a complex WM span task. The authors report more transfer to an untrained complex span task for the group with higher PI-demands in the training task, compared to a control group with lower PI-demands. Further, the training group with more PI in the training task experienced fewer intrusions in an untrained thought-monitoring task than the control group (Bomyea and Amir, 2011). As the study did not focus on the specific processes underlying WM training and transfer, the authors did not report whether PI was actually reduced in the training tasks at the end of the training. Taken together, there is still little evidence on whether resistance to PI can actually be improved by a specific WM training, and, as of yet, there is no study investigating this issue in older adults as a main target population for cognitive interventions.

Despite the retraction of the study of Persson and Reuter-Lorenz (2008), the idea to manipulate demands on PI resistance within WM training remains intriguing. Furthermore, to our knowledge, no previous attempts exist to train resistance to PI in older adults, although an improvement of this ability and a possible generalization to other tasks would have significant implications for treatments in healthy and pathological aging. In particular, older adults have been found to be more susceptible to PI in different WM tasks compared to younger adults (cf. May *et al.*, 1999; Jonides *et al.*, 2000; Schmiedek *et al.*, 2009; Loosli *et al.*, 2014). Hasher *et al.* (2002) showed that compared to young adults, older adults are less able to delete no longer relevant items from WM due to deficient inhibition processes, which results in the inability to differ between relevant and irrelevant items and therefore in more intrusion errors, i.e. more PI.

In the present study, we adapted the design of Persson and Reuter-Lorenz (2008) and trained two groups of old adults with a recent-probes task and an *n*-back task over the course of two weeks. Both tasks have been previously used to measure resistance to PI within WM in the elderly (Loosli *et al.*, 2014). As in the original training study of Persson and Reuter-Lorenz (2008), the only difference between the two training groups was the amount of PI-trials in the training tasks (high vs. low PI). The training tasks of the high-PI group contained recent trials that were closer to the probe than those in the low-PI group, and should therefore elicit more PI. The low-PI group was expected to show less or no PI, as their recent trials were at more distant positions (see also Hartshorne, 2008; McCabe and Hartman, 2008). We further investigated *near transfer*, which is defined as transfer to tasks that involve similar cognitive processes in the outcome measures as the trained tasks, i.e. interference resolution. As in the study of Persson and Reuter-Lorenz (2008), we used a verb generation and a paired-associates task, and further added a Stroop task to measure transfer to prepotent response inhibition. Performance in this task has also been shown to be related to WM capacity (Kane and Engle, 2003) and to be improved in older adults following a WM training (Borella *et al.*, 2010). Moreover, we extended the original study by adding measures of *far transfer* that is defined as an improvement in processes that are conceptually not directly related to trained tasks such as gF. Although gF is related to WM capacity (e.g. Kyllonen and Christal, 1990), results regarding transfer effects on gF after WM training are contradictory (cf. Melby-Lervåg and Hulme, 2013; Redick *et al.*, 2013). However, gF is also related to resistance to PI (Gray *et al.*, 2003; Bunting, 2006; but see also Emery *et al.*, 2008), and behavioral and neural correlates of PI were even found to explain a significant proportion of common variance between WM span and gF (Burgess *et al.*, 2011). Furthermore, interference resolution has been suggested to be an important cognitive process behind transfer effects on gF following WM training with *n*-back tasks (Jaeggi *et al.*, 2014). It is hence of high interest whether a WM training varying the amount of PI would lead to differential improvement in gF.

Taken together, we addressed the following three major research questions:

(i) *Can resistance to PI be trained in older adults by a WM training with increased PI demands?* We hypothesized that the group with the high-PI condition in the training would show more improvement regarding resistance to PI from the first to the last training day in both training tasks compared to the group with low-PI demands.

(ii) *Does a possible improvement in resistance to PI generalize to other interference tasks (near transfer) and will this improvement be larger than a possible transfer of a WM training with less PI-demands?* Given that all near-transfer tasks required interference resolution processes (Thompson-Schill *et al.*, 1997; Milham *et al.*, 2001; Henson *et al.*, 2002) and an improvement of PI resistance in the training tasks was expected, we postulated that this improvement would generalize and lead to an improvement in the interference measures of the near-transfer tasks. This improvement was expected to be higher in the high-PI group compared to the low-PI group, as it was specifically trained.

(iii) *Will a WM training with higher PI-demands lead to more far transfer to gF and processing speed than a WM training with less PI?* Regarding gF, we hypothesized that training with high-PI demands will lead to more improvement than an identical training with low-PI demands, as resistance to PI has been shown to mediate the relationship between WM and gF (Burgess *et al.*, 2011). Concerning processing speed, we expected that both groups would show identical improvement, that is, we expected both groups to have shorter RTs after training. Therefore, this task served, besides measuring far transfer, as a control variable.

To obtain valid data, we controlled for potential confounds using the following approach: First, by including an active control group with low PI-demands, we controlled for effects that might also lead to an improvement, e.g. expectancy effects or effects of meeting in a group on a daily basis. Second, to reduce any pre-existing differences in cognitive functions or other individual characteristics, we matched the participants before assigning them randomly to the training conditions (quasi-random design). Third, the study was conducted double-blind, i.e. neither the participants nor the experimenters knew a participants' assignment to a specific experimental condition.

Methods

Participants

Participants were recruited from the lab's participant database as well as with flyers at public places. Inclusion criteria were age of sixty years or older and availability in terms of time to take part (i) in five daily training sessions per week for two consecutive weeks as well as (ii) in the pre-, post- and follow-up assessments. During training, participants were allowed to be absent two days at maximum, but not on Day 1 or Day 10. Exclusion criteria were (i) psychiatric or neurological disorders or the use of psychotropic medication. After data acquisition, (ii) an incomplete data set due to technical or motivational reasons, or (iii) extreme outliers in the data would also lead to exclusion. All participants gave written informed consent prior to participation. The study protocol was approved by the local ethics committee. Participants who finished the study received 50€, those who quit before the follow-up assessment received proportionate compensation.

A total of 39 participants were enrolled for participation. Exclusion criteria and the process of inclusion and exclusion are described in the supplementary material (available as supplementary material attached to the electronic version of this paper at www.journals.cambridge.org/10.1017/S1041610215001519). The final sample consisted of 25 participants: 14 participants in the high-PI training group (mean age = 68.9 years, *SD* = 5.5, 9 women) and 11 in the low-PI group (mean age = 68.6 years, *SD* = 5.7, 9 women). Groups did not differ significantly regarding age ($t(23) = 0.17$, $p = 0.867$). Participants completed on average 14.8 years of education (*SD* = 4.3, no difference between groups, $p = 0.564$). Additionally, the short form of the Geriatric Depression Scale (GDS; Yesavage and Sheikh, 1986) was completed as screening instrument for depression. All included participants scored in the normal or low borderline range at the pre-test and follow-up session, respectively (score range from 0 to 6). Groups did not differ significantly at pre- ($t(23) = 0.62$, $p = 0.544$) or follow-up session ($t(23) = 0.70$, $p = 0.489$). See also online supplementary material, Section S1, for a more detailed description of the sample.

Experimental tasks and design

Training tasks

Recent-probes task: The recent-probes task (cf. Monsell, 1978) has previously been used to assess PI-processes within WM (see Jonides and Nee, 2006; Unsworth, 2010) and was adapted from Loosli *et al.* (2014). Twelve animal pictures (non-colored line drawings) from the Snodgrass and Vanderwart object collection (Snodgrass and Vanderwart, 1980) were used as stimuli (see Figure 1A). In each trial, four pictures were displayed simultaneously as a target for 2,000 ms in a 2 × 2 array on a computer screen. These pictures had to be encoded and then maintained over a delay of 2,500 ms during which a fixation cross was shown. Then, a single animal was presented centrally as a probe for 1,500 ms and the task was to evaluate whether this probe matched one of the targets presented before or not (yes/no). If the probe was part of the target set in the present trial, the left button of a computer mouse had to be pressed (positive trial) and if not, the right button had to be pressed (negative trial). Participants were instructed to answer as quickly and accurately as possible. The task was set up in the Python programming language (http://www.python.org/).

The task comprised a 2 × 2 design: The factor *recency* reflected whether the probe was part of the target set in the previous trial (recent) or not (non-recent); the factor *probe type* indicated the kind of response that had to be given (positive or negative; i.e. whether the probe matched one of the items in the target set in the current trial or not). Recent-negative trials should elicit PI, and participants with lower resistance to PI were expected to erroneously respond with "yes" instead of "no" more often (i.e. they were expected to show more false-positives) and to take longer for this decision. Participants were not aware of the manipulation of the factor recency. In total, there were 64 trials in each run. Accuracy and RTs were recorded and PI was calculated as the difference between recent and non-recent negative trials.

As in the original study of Persson and Reuter-Lorenz (2008), two different versions of this task were used: The high-PI training group was trained and assessed with the task as described above (PI came from the most recent trial, i.e. one trial back). The low-PI group was assessed with the high-PI training task on Days 1 and 10 (so as to compare PI susceptibility between groups), whereas a low-PI version was applied on Days 2 to 9. In the low-PI version, PI came from the target set two trials back and therefore was expected to elicit substantially smaller PI effects. In this respect, it was previously shown that the further back the recent probe occurred as target, the less PI was caused in the current trial (Hartshorne, 2008).

N-back task: The visual two-back task was adapted from Loosli *et al.* (2014). Twelve pictures of daily objects (non-colored line-drawings, from Cycowicz *et al.*, 1997; see also Snodgrass and Vanderwart, 1980) were used as stimuli (see Figure 1B). Pictures were presented serially for 2,000 ms each, one at a time, on the computer screen. Picture presentations constituted individual trials that were separated by the presentation of a fixation cross for 2,500 ms. Participants had to indicate whether the object in the current trial matched the one presented two trials before. If this was the case, the left button of a computer mouse had to be pressed (positive); else, the right button had to be pressed (negative). Participants were instructed to answer as quickly and accurately as possible. The task was set up in Python (see also above).

There were three different conditions: *Targets,* which were the same stimuli as two trials before, they required a "yes" answer. *Lure* trials comprised stimuli already presented three trials before and required a "no" answer. In trials with *nones*, the object was presented longer than six trials before hence requiring a "no" answer. Lure trials should elicit PI, and participants with lower resistance to PI were expected to erroneously respond with "yes" instead of "no" more often and to take longer for this decision. Participants were not aware that there were lure trials. In total, the *n*-back consisted of 100 trials, 20 of which were targets (20%) and 20 were

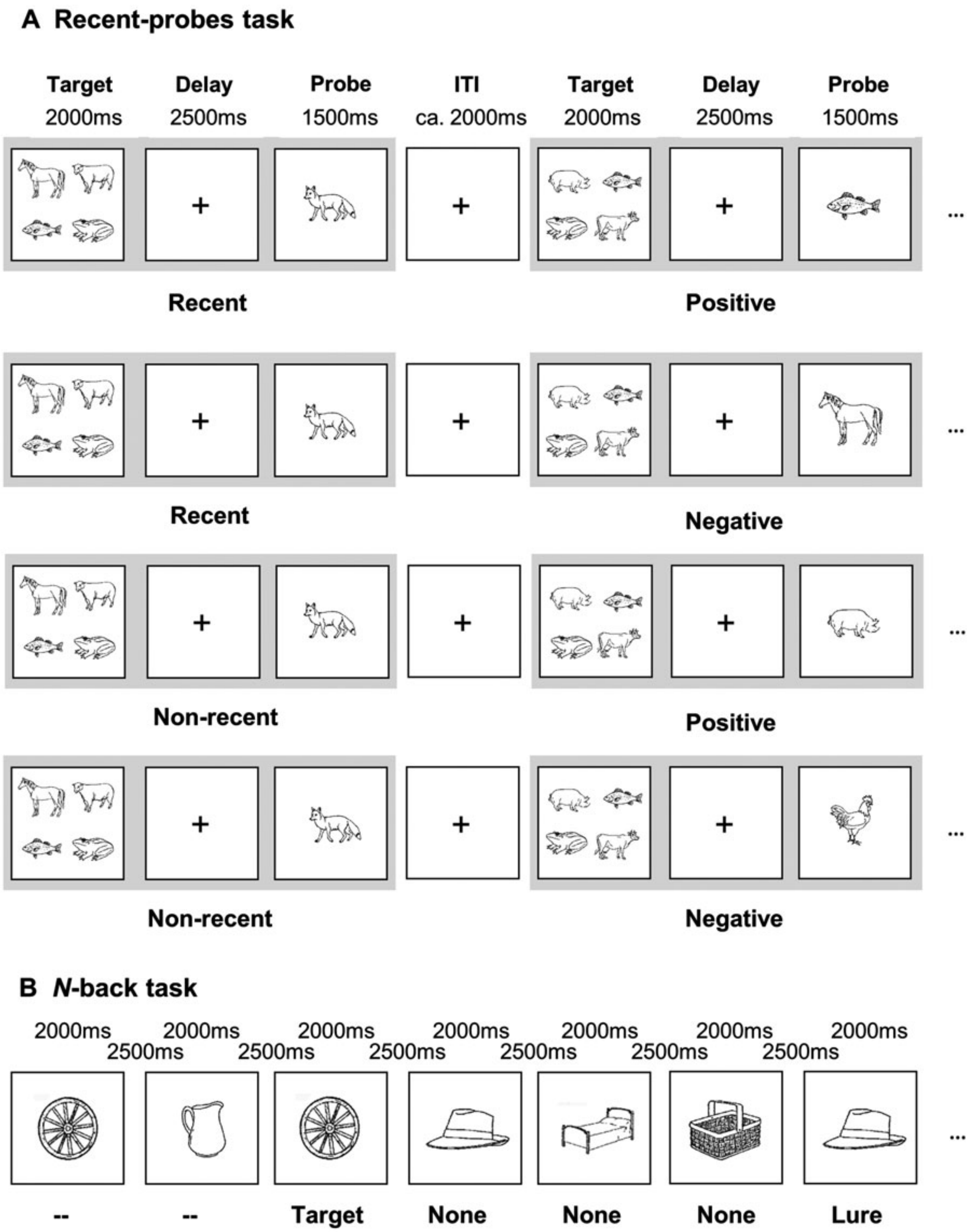

Figure 1. Training tasks. In the recent-probes (A), a target set of four animals had to be encoded, to be held in mind, and when a probe with a single animal appeared, participants had to decide whether it was part of the four animals shown before (left mouse button) or not (right button). There were four conditions, the recent-negative condition elicited PI (cf. Loosli *et al*., 2014). In the *n*-back (B), participants had to press the left button if the present object was the same as two positions before; else, the right button had to be pressed. There were three conditions: targets, lures (PI) and nones (cf. Loosli *et al*., 2014). On Days 1 and 10, both training groups trained with the high-PI conditions as displayed here. On Days 2 to 9, the high-PI training group used the same tasks, while the low-PI group had identical tasks with less (recent-probes) or no (*n*-back) PI.

lures (20%). As in the recent-probes task, accuracy and RTs were recorded and PI was calculated as difference between lures and nones.

Again, two different versions of this task were used. The high-PI training group trained with the task as described above, i.e. with three different conditions. The low-PI group was assessed on this task on Days 1 and 10, whereas a low-PI version without lures (hence consisting of 20 targets and 80 nones) was applied on Days 2 to 9. It was previously shown in a 2-back task that 3-back lures were answered with significantly more errors than nones and also significantly worse than 4-, 5-, or 6-back lures (McCabe and Hartman, 2008).

Pre- and post-test tasks on near and far transfer

For assessing near transfer, we used a verb-generation and a paired-associates task, as in the original study of Persson and Reuter-Lorenz (2008), but further added a Stroop task. In these three tasks, resolution of interference is required by selecting a correct response among different competing alternatives (Thompson-Schill *et al.*, 1997; Milham *et al.*, 2001; Henson *et al.*, 2002; Persson *et al.*, 2007). For assessing far transfer, a digit-symbol substitution test was used to investigate unspecific transfer to processing speed and a matrix reasoning test was used as a measure of gF. More detailed information can be found in the online supplementary material, Section S2.

General procedure

Enrolled participants were matched regarding age and sex, and, if available from a previous study of our lab in which some of the participants took part (Köstering *et al.*, 2014), WM performance (*n*-back and recent-probes task), speed, and gF. After matching, they were assigned to the high- or low-PI training condition. The study design was double-blind, i.e. experimenters were not aware of the group membership of the participants, and participants neither knew that there were two different training conditions nor which group they belonged to. The training was conducted in a quiet computer lab in groups. There were ten training days (two weeks, Mondays to Fridays). On Day 1 and Day 10, all participants accomplished a high-PI version of the tasks, as these days served as test sessions for the training tasks. On Days 2 to 9, participants received either the high- or the low-PI versions, dependent on their group membership. Task difficulty was constant during all runs. The total amount of time spent on the training tasks was about 32 min per day. Further information about the matching process and the training can be found in the supplementary material, Section S3.

Pre-test sessions were conducted on the weekend before the training started; post-tests on the weekend directly after the second week of training. The follow-up session, which tested for possible long-term transfer, was conducted eight to nine weeks after the post-test session ($M = 8.51$ weeks, $SD = 0.29$, range = 8.14–9.29 weeks).

Data and outlier analysis

In both training tasks, trials without responses were discarded. Additionally, the first two trials of the *n*-back were excluded, as no picture had appeared two positions back for these trials. RTs were aggregated across all included trials. However, to control for extreme outliers in the RT data, RTs for single trials were z-transformed, the reference for this transformation was the individual participant, training run, task, and condition. Trials with $z > 3$ or < -3 were excluded from further processing. In the recent-probes task, 3.5% of all trials were excluded at Day 1 and 1.5% at Day 10. In the *n*-back, 5.4% and 2.8% were excluded on the first or the last day of training, respectively. The remaining trials were then aggregated, and median RT and mean accuracy were calculated for each condition and participant.

For both training tasks, sensitivity (d') was calculated according to Stanislaw and Todorov (1999) as a measure of WM performance. Sensitivity was used, as the proportion of corrects can be biased by tendencies towards "yes" or "no" answers. A higher value in d' reflects better ability to discriminate between positive and negative trials (Stanislaw and Todorov, 1999). These calculations were based on non-recent trials (recent-probes) or on targets and nones (*n*-back). Interference scores for the training tasks were calculated as the difference between interference and non-interference trials (only from negative trials).

On the participant level, d' and PI-scores of the training tasks on Day 1, and data of the transfer tasks at pre-test were again checked for outliers ($z <> \pm 3$ SD). However, as the performance of all participants was within this range, no data were excluded.

For information on outlier analysis in the transfer tasks, see online supplementary material, Section S4.

Data were analyzed with IBM SPSS Statistics, release 20.0.0 (IBM Corp., Chicago, IL). Statistic tests were based on a significance level of $\alpha = 0.05$. Training and transfer effects were analyzed with repeated-measurements analyses of variance (ANOVAs). Cohen's d and η_p^2 were calculated as measures of effect size (Cohen, 1988; see also Durlak, 2009). As the commonly used formula

Table 1. Descriptives, effect sizes and F-values for performance changes in the training tasks

TASK	MEASURE	GROUP	DAY 1		DAY 10		COHEN'S *D*	*F* TIME	*F* GROUP	*F* GROUP × TIME
			MEAN	*SD*	*MEAN*	*SD*				
Recent-probes										
	d'	High-PI	2.26	0.74	2.80	0.58	0.79	11.833**	1.973	0.031
		Low-PI	1.83	1.11	2.43	0.84	0.59			
	RTs (ms)	High-PI	1,284	177	1,137	118	0.98	53.715**	0.060	5.544*
		Low-PI	1,371	291	1,084	142	1.25			
	PI Accuracy	High-PI	0.13	0.12	0.06	0.07	0.72	4.559*	1.231	0.419
		Low-PI	0.09	0.07	0.05	0.08	0.48			
	PI RT-costs	High-PI	84	102	105	131	− 0.17	1.934	0.000	0.269
		Low-PI	71	102	117	74	− 0.50			
N-back										
	d'	High-PI	2.81	0.77	3.60	0.64	1.08	18.631***	0.044	0.127
		Low-PI	2.79	1.00	3.72	0.68	1.05			
	RTs (ms)	High-PI	1,275	237	1,084	213	0.85	16.287***	0.035	2.562
		Low-PI	1,421	561	977	195	1.06			
	PI Accuracy	High-PI	0.37	0.18	0.16	0.17	1.21	47.978***	0.168	0.398
		Low-PI	0.38	0.15	0.20	0.19	1.02			
	PI RT-costs	High-PI	149	170	139	76	0.08	0.177	0.018	0.480
		Low-PI	130	137	169	133	− 0.28			

Note*:* *SD* = standard deviation; d' = d-prime; RT = reaction time; ms = milliseconds; PI = proactive interference; d' and RTs comprised only trials without proactive interference and non-recent trials. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

may overestimate effects in small samples (Durlak, 2009), Cohen's d was corrected for small sample sizes (Hedges and Olkin, 1985). According to Cohen (1988), $d > 0.2$ represents a small effect, $d > 0.5$ a medium effect and $d > 0.8$ a large effect.

Results

Two-sample *t*-tests showed that there were neither significant differences between both training groups regarding pre-test scores of the transfer measures for near (all $p > 0.181$) and far transfer (all $p > 0.606$) nor regarding PI-scores or WM performance (d') in the training tasks on the first training day (all $p >$ 0.284).

In terms of a manipulation check, paired-sample *t*-tests were used to verify the presence of interference effects in both training tasks as well as in the near-transfer tasks in the overall sample before training. In the recent-probes task, recent negatives were answered less accurate, $t(24) = -5.34$, $p < 0.001$, and more slowly, $t(24) = 3.89$, $p < 0.001$, than non-recent negatives. In the *n*-back task, similarly, accuracy was lower in lures than in nones, $t(24) = -11.50$, $p < 0.001$, and lures were answered slower than nones, $t(24) = 4.57$, $p < 0.001$. Therefore, PI-effects were present in accuracy and RT in both tasks, which was a pre-condition for this training of resistance to PI.

Training effects

Training of WM performance

Accuracy: Improvement of WM performance irrespective of the PI-manipulation was analyzed with univariate ANOVAs (two for global improvement of WM performance (d') in both tasks, and two for global improvements in RTs in both tasks). Only non-recent trials were included in this analysis in order to prevent an interference bias from recent trials.

As displayed in Table 1, there were significant main effects of time in both tasks (recent-probes: $F(1, 23) = 11.83$, $p < 0.01$, $\eta_p^2 = 0.34$; *n*-back: $F(1, 23) = 18.63$, $p < 0.001$, $\eta_p^2 = 0.45$), indicating that WM performance increased from Day 1 to Day 10. However, the increase was similar for both training groups and thus independent of particular demands on PI resistance (i.e. no main effects of group, recent-probes: $F(1, 23) = 1.97$, $p = 0.174$, $\eta_p^2 = 0.08$; *n*-back: $F(1, 23) = 0.04$, $p = 836$, $\eta_p^2 = 0.00$, and no interactions between time and group, recent-probes: $F(1, 23) = 0.03$, $p = 0.863$, $\eta_p^2 = 0.00$; *n*-back: $F(1, 23) = 0.13$, $p = 724$, $\eta_p^2 = 0.01$).

RT: We further tested whether there was a global decrement regarding RTs of the non-recent trials. In the recent-probes task, both groups were faster on Day 10 than on Day 1 (main effect of time, $F(1, 23) = 53.72$, $p < 0.001$, $\eta_p^2 = 0.70$), and this effect was stronger in the low-PI group than in the high-PI group (Time × Group interaction, $F(1, 23) = 5.54$,

$p < 0.05$, $\eta_p^2 = 0.19$). There was no main effect of group, $F(1, 23) = 0.06$, $p = 0.808$, $\eta_p^2 = 0.00$. Regarding the *n*-back task, both groups improved identically in terms of RT, as there was a main effect of time, $F(1, 23) = 16.29$, $p < 0.001$, $\eta_p^2 = 0.42$, but no interaction, $F(1, 23) = 2.56$, $p = 0.123$, $\eta_p^2 = 0.10$ and also no main effect of group, $F(1, 23) = 0.04$, $p = 0.854$, $\eta_p^2 = 0.00$.

To summarize, WM performance improved in both tasks in the course of the training, and the amount of improvement was similar for both groups. RTs were faster in both tasks and both groups after training, this improvement was slightly stronger for the low-PI group. An overview on the descriptive and inferential statistics is provided in Table 1.

Training of resistance to PI

Improvement of resistance to PI was analyzed with univariate ANOVAs for accuracy and RTs separately in both tasks. Descriptives for Days 1 and 10 are reported in Table 1. Performance in PI-measures across all training runs from Days 1 to 10 is displayed in Figure 2, the respective changes in PI-scores are displayed in Figure 3. As is evident from Figure 2, the low-PI training group showed across Days 2 to 9 almost constantly less PI in the recent-probes task than the high-PI group, as PI came not from the most recent trial, but from two trials back, and therefore was expected to be substantially weaker. This difference (average PI-effect from Day 2 to Day 9) was statistically significant for errors ($t(18.96) = 3.019$, $p < 0.01$) and for PI-related RT-costs ($t(23) = 2.83$, $p < 0.05$).

Accuracy: Univariate ANOVAs showed that PI-scores decreased over time in both tasks (recent-probes: $F(1, 23) = 4.56$, $p < 0.05$, $\eta_p^2 = 0.17$; *n*-back: $F(1, 23) = 47.98$, $p < 0.001$, $\eta_p^2 = 0.68$, see also Figure 2 and Table 1). However, there were no differential effects regarding this improvement, as there were no significant Time × Group interactions (recent-probes: $F(1, 23) = 0.42$, $p = 0.524$, $\eta_p^2 = 0.02$; *n*-back: $F(1, 23) = 0.40$, $p = 0.535$, $\eta_p^2 = 0.02$) as well as no main effects of group (recent-probes: $F(1, 23) = 1.23$, $p = 0.279$, $\eta_p^2 = 0.05$; *n*-back: $F(1, 23) = 0.17$, $p = 0.685$, $\eta_p^2 = 0.01$).

RT: Univariate ANOVAs regarding both tasks indicated no improvement in PI-related RT-costs (recent-probes: $F(1, 23) = 1.93$, $p = 0.178$, $\eta_p^2 = 0.08$; *n*-back: $F(1, 23) = 0.18$, $p = 678$, $\eta_p^2 = 0.01$). As can be seen in Figure 2B, PI-related RT-costs were fluctuating +/− 50ms around the mean and did not decrease across training runs. Further, there were no main effects of group (recent-probes: $F(1, 23) = 0.00$, $p = 0.989$, $\eta_p^2 = 0.00$; *n*-back: $F(1, 23) = 0.18$, $p = 895$, $\eta_p^2 = 0.01$) and no interactions (recent-probes: $F(1, 23) = 0.27$, $p = 0.609$, $\eta_p^2 = 0.01$; *n*-back: $F(1, 23) = 0.48$, $p = 495$, $\eta_p^2 = 0.02$).

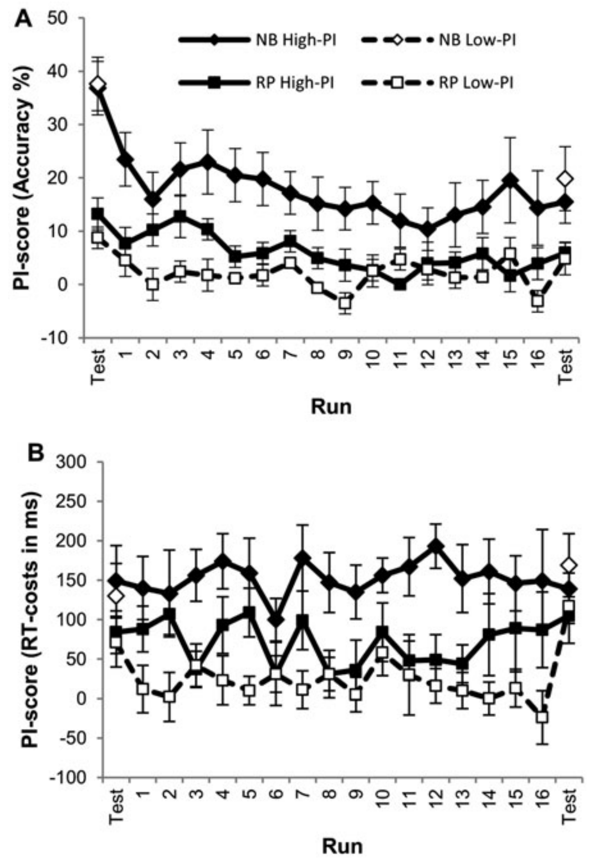

Figure 2. Change in proactive interference (PI) across training runs. Mean and standard errors of PI-scores (difference between interference and non-interference negative trials) for both tasks and groups regarding accuracy (2A) and reaction-time (RT) costs (2B). Displayed are individual training runs, which might not necessarily have occurred on the same day for all participants, as some were absent on some days (at maximum two days missing were allowed). Under these circumstances, the *N* for the earlier runs is higher than the *N* in the later runs. The *n*-back for Days 2 to 9 for the low-PI group contained no PI, therefore, only values for Days 1 and 10 are displayed for this group. NB = *n*-back, RP = recent-probes.

In sum, both training groups could improve resistance to PI in both tasks in terms of accuracy, however, there was no difference between both groups' improvement. There was no reduction in PI-related RT-costs in either of the groups or tasks.

Immediate and long-term effects of near and far transfer

Before training, high-interference conditions were answered more slowly than low-interference conditions in all near-transfer tasks (verb generation: $t(24) = 7.45$, $p < 0.001$; paired associates:

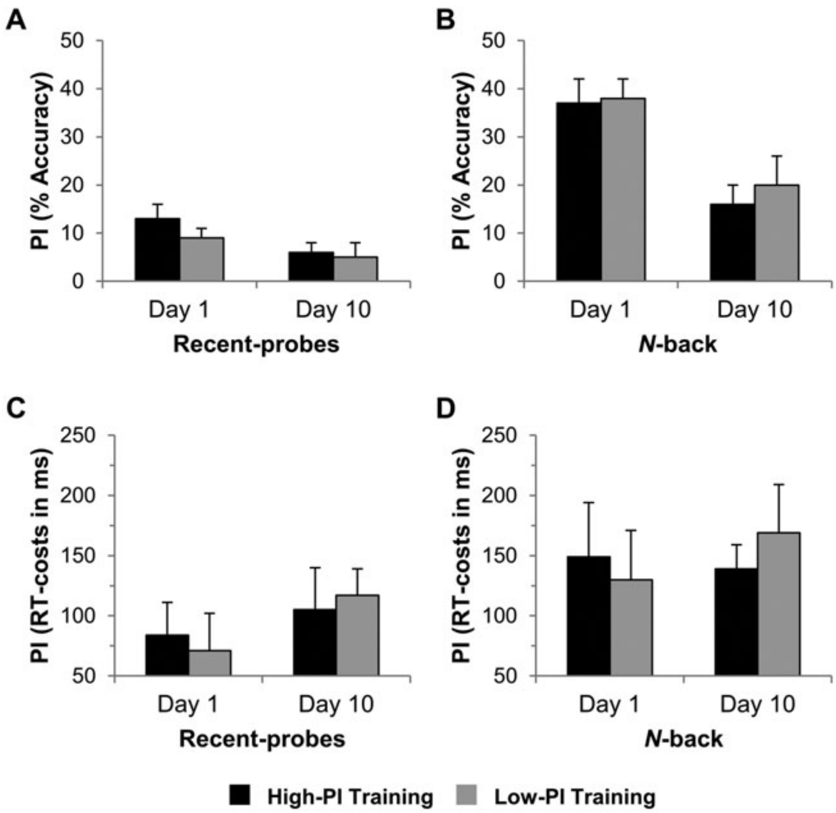

Figure 3. Proactive interference (PI) scores on the first and last training day. Mean and standard errors of the PI-scores for accuracy (A–B) and reaction times (RT, C–D) for both training groups on the first and last training day.

$t(22) = 6.15$, $p < 0.001$; Stroop test: $t(24) = 11.41$, $p < 0.001$). Performance in the transfer tasks at pre-test, post-test and follow-up sessions for both groups is displayed in Figure 4 and reported in Table S1 in the supplementary material.

Separate repeated-measurements ANOVAs with training group (high- vs. low-PI) as between-participants factor and time (pre-test vs. post-test) as within-participant factor were conducted for the three near-transfer tasks as well as for the two far-transfer tasks. There were no differential near- or far-transfer effects, neither directly after the training nor eight weeks later (Time × Group interactions all $p \geq 0.335$). Results of the ANOVAs can be found in the supplementary material (Sections S5 and S6).

Discussion

The overall aim of the present study was to investigate whether resistance to PI can be effectively improved by a WM training. Two groups of older adults were trained with identical WM tasks over the course of two weeks, with the only difference that PI demands were experimentally varied between groups (high vs. low PI). Another main objective was to investigate whether a potential improvement in resistance to PI would transfer to other interference tasks (near transfer) or to gF and processing speed (far transfer). In the next sections, the results will be discussed.

Training effects

Results revealed a global improvement in WM performance in both the recent-probes and the *n*-back task that was similarly pronounced in both training groups (see Table 1). Such training effects have already been shown in other studies with older adults (e.g. Li *et al.*, 2008; Dahlin *et al.*, 2008b) and demonstrate still existing cognitive plasticity in this age group.

In addition, PI-related errors were reduced not only in the high-PI but also in the low-PI group (see Table 1 and Figures 2 and 3). To our knowledge, this is the first study demonstrating generally reduced PI-susceptibility following WM training in old adults. But contrary to our expectations, the improvement was not larger in the high-PI group. Thus, the WM training itself and the resulting higher WM performance at the end of the training in both groups might have led to better resistance to PI, regardless whether resistance to PI was explicitly trained or not. As a reciprocal relationship between resistance to PI and WM capacity can be assumed (see e.g. May *et al.*, 1999; Kane and Engle,

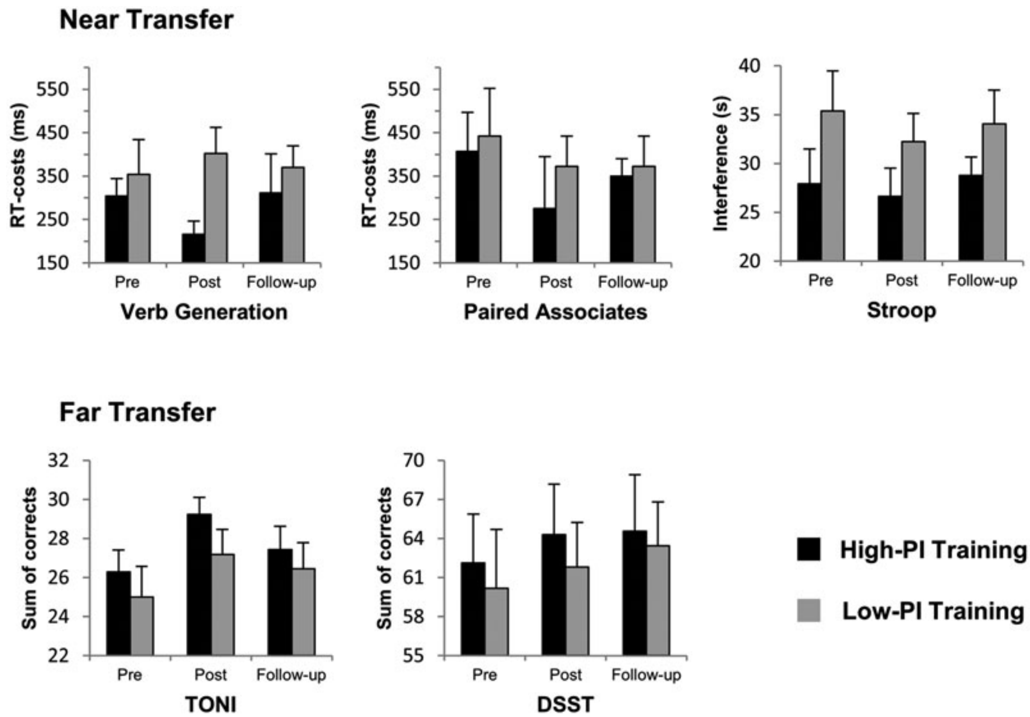

Figure 4. Short- and long-term changes in transfer measures. Mean and standard errors of performance for both training groups. RT = reaction time, PI = proactive interference, TONI = Test of Nonverbal Intelligence, DSST = Digit Symbol Substitution Test.

2000; Mecklinger *et al.*, 2003), both functions could possibly influence each other. But improved resistance to PI was only found in terms of a decrement in PI-related errors, whereas a similar decrease in PI-related RT-costs was not observed in either of the groups (see Table 1 and Figures 2 and 3). Individual differences in gain on WM performance or resistance to PI were barely related to cognitive performance as assessed previous to the intervention. Further, the training effects should be interpreted with caution, as no passive control was used. Unspecific effects such as re-testing, expectancy effects or meeting in a group on a daily-basis (cf. Willis, 2001) could also be a source of this general improvement.

The question remains why the high-PI training group could not show stronger benefits in PI resistance than the low-PI group. Although the relevant interaction effects for corroborating a differential improvement were not significant, effect sizes regarding the improvement in PI-related errors were higher in the high-PI group than in the low-PI group (see Table 1). As the *N* in our study was rather small, one could argue that low power might be responsible for the non-significant interactions. However, effect-sizes for the interactions were only small (see Result section), and sample sizes of 80 (*n*-back PI-accuracy) or about 200 (other PI-measures, see Table S2 in Section S7 in the online supplementary material) would have been required to obtain significant differential effects. That is, even if potentially rendered significant in a larger sample, potential effects of a differentially improved PI resistance following a specific PI training are likely to be too small to become practically relevant.

Another methodological aspect may concern the non-adaptivity of the applied WM training tasks. That is, the difficulty of the tasks in terms of WM load and the amount of PI remained constant across the training and were not adaptively adjusted to the participants' performance. Some participants were on ceiling level already in the first days of the training, whereas others struggled with the demands imposed by the tasks until the end of the second training week. It may hence be speculated that the gain in PI resistance could have been larger using an adaptive training approach or by solely focusing on participants with a compromised WM performance and/or PI susceptibility.

Near transfer

Neither differential nor global short- or long-term improvement in non-trained tasks was observed. From a theoretical perspective it can be argued that, as there was no differential improvement in the training tasks regarding PI, no differential improvement in the transfer tasks could have been expected (see Willis, 2001). But as both groups showed improved PI-accuracy scores and also ameliorated WM performance, the absence of a general improvement in the interference scores of the near-transfer tasks remains to be clarified.

Several approaches have previously been discussed to explain transfer from trained to non-trained tasks, e.g. structural similarities between training and transfer tasks (Salomon and Perkins, 1989), overlapping cognitive processes between tasks (Dahlin *et al.*, 2008b; Shipstead *et al.*, 2010), or overlapping neural correlates of the tasks (Dahlin *et al.*, 2008b). From a perspective emphasizing structural similarities between tasks as a prerequisite for transfer, one could argue that training and near-transfer tasks were not directly comparable. The paired-associates task is structurally the most similar transfer task when compared to the training tasks, as it also involves encoding and maintenance of information. However, compared to the training tasks, larger amounts of information had to be encoded and maintained in long-term rather than short-term memory. In addition, assessment of memory performance (and interference effects) was based on free recall and not on recognition as in the applied WM training tasks. That is, similarity of task structure was only given on a superficial level. The task structure of the other near-transfer tasks was completely different from the WM training tasks. Further, while the WM training tasks consisted of visual stimuli, verbal material was used in the near-transfer tasks. Taken together, it may hence be possible that there was no near transfer because of the lacking structural similarities of training and transfer tasks.

However, with respect to underlying cognitive processes and neural correlates, all near-transfer tasks as well as the training tasks required selecting a correct answer among competing alternatives, and therefore the resolution of interference (e.g. Thompson Schill *et al.*, 1997; Milham *et al.*, 2001; Henson *et al.*, 2002). Resolution of PI in the recent-probes task, in verb generation and in paired associates has previously been associated with activations in the prefrontal cortex, namely in the left inferior frontal gyrus (LIFG, Thompson-Schill *et al.*, 1997; Henson *et al.*, 2002; Nelson *et al.*, 2009). It is hence plausible to assume that training-related improvements in one task might spill over to a performance improvement in other tasks that recruit the same brain areas, but were not trained (see e.g. Dahlin *et al.*, 2008b). Persson *et al.* (2007; 2013) even showed negative transfer on interference resolution (verb generation and paired associates) after a very short (single-session) training with a letter recent-probes task in young adults. That is, a short, intensive training reduced the ability to resolve interference in tasks which, compared to the training tasks, required comparable processes or involved overlapping neural correlates (Persson *et al.*, 2007). These results were explained as process-specific fatigue effects between interference tasks that share similar processes and neural correlates. Taken together, neuroimaging studies and studies on negative transfer indicate a relationship between several interference-resolution tasks that were also used in our training study.

However, the similarities in underlying cognitive processes and neural correlates may have been overridden by the differences between tasks and may hence have been too weak to serve as basis for effective transfer. For example, interference in terms of actual PI only occurred in the paired associates and not in verb generation or in the Stroop task, but was the critical process to be trained. Further, we are not aware of previous training studies using the verb generation and the paired-associates task as transfer task – besides the two studies on negative transfer cited above (Persson *et al.*, 2007; 2013) – so the present results cannot be compared with other work directly. However, WM training studies that used the Stroop task as transfer measure revealed mixed results, although a meta-analysis showed an overall small to moderate effect (Melby-Lervåg and Hulme, 2013). Likewise, Borella *et al.* (2010) were able to show transfer to the Stroop task after WM training in older adults. Their training used an adaptive procedure and a greater variation of task demands than the present study, which might possibly have been more advantageous to achieve transfer. Other methodological reasons that could account for the absence of near-transfer effects and more importantly, also for the total absence of global improvement in these tasks despite that all participants received a training of WM, could be on the one hand the relative shortness of the present intervention compared to other training studies in the elderly (see Richmond *et al.*, 2011, who list five training studies with training durations between 675 and 6,060 min; in contrast, our participants only trained for 32 min a day for two weeks). On the other hand, small power due to small sample size might provide an explanation for the lacking global and differential transfer effects. As shown in Table S2 in the online supplementary material, more than 100 participants would have been needed to obtain significant interaction effects in the near-transfer tasks, again questioning the practical relevance.

Far transfer

The matrix reasoning task used in this study shares neither structural similarities with one of the WM training tasks (i.e. similarities between task structure, task presentation or task conduction) nor conceptual similarities, as the training tasks and the gF task rely on different theoretical constructs and represent different cognitive functions. The absence of these similarities is indeed a criterion for defining far transfer tasks, but gF has nonetheless been shown to be related to PI (Bunting, 2006; Burgess *et al.*, 2011; but see also Borella *et al.*, 2006; Emery *et al.*, 2008). Both WM capacity and gF seem to share interference resolution as common underlying process, although the tasks differ considerably in their appearance. However, in our study, the high-PI training group did not show significantly larger improvement regarding gF when compared to the low-PI group, although the estimated effect size was considerably larger in the high-PI training group than in the low-PI training group (see Table S1 in the online supplementary material). Nonetheless, the strength of the interaction effect was small, as also confirmed by post-hoc power-analyses (see Table S2 in the online supplementary material). More than a hundred participants would have been needed to obtain a significant interaction effect regarding gF (see Table S2), which is considerably more than the usual sample size in training studies (see e.g. Melby-Lervåg and Hulme, 2013, for an overview) and beyond any practical relevance. Nevertheless, there was a global improvement in the gF task, independent of group assignment, which was still marginally significant at the follow-up session eight weeks after the training. This outcome replicates findings from previous WM training studies that also showed improvements in gF after WM training in older adults (e.g. Borella *et al.*, 2010; Carretti *et al.*, 2013b; Zinke *et al.*, 2014). However, whether this unspecific effect is a consequence of the WM training or whether it is only a re-test effect cannot be determined for the present data, as no passive control was used. Oelhafen *et al.* (2013) also found an unspecific effect on gF after a WM training with differing interference control demands in young adults, however, when compared against a passive control group, the interaction effect did not reach significance. There is an ongoing debate in the literature regarding whether gF can be improved after WM training. A recent meta-analysis showed that effects of WM training on non-verbal ability were small (Melby-Lervåg and Hulme, 2013; for a more thorough discussion of transfer to gF after WM training, see also Harrison *et al.*, 2013; Redick *et al.*, 2013; Jaeggi *et al.*, 2014). Also in older adults, transfer to gF has not consistently been found (e.g. Richmond *et al.*, 2011; von Bastian *et al.*, 2013).

There was no transfer on processing speed directly after training, but there was a marginal effect from pre-test to follow-up measurements. The large general decrements in RTs in both WM training tasks (see Table S1 in the online supplementary material) might have led to a general improvement in processing speed in both groups, even though the improvement was not significant directly after training. Dahlin *et al.* (2008a) similarly found marginally improved performance on the DSST in older adults after a computerized updating training compared to a passive control group. However, performance decreased below pre-test level at the follow-up measurement in their participants. In the present study, although there was a global effect, there was no differential transfer, although the low-PI group improved more in global RTs in the training tasks. As in the other measures, the result of the general improvement should not be over-interpreted, as no passive control was used.

Conclusions

Implementing increased PI-demands in WM training tasks did not improve resistance to PI more effectively than a WM training with less PI in older adults, which might have numerous reasons, as discussed above. But irrespective of the unspecificity of the experimental PI manipulation, the present results demonstrate for the first time that resistance to PI can be reduced in older adults by a short, repetitive WM training as such. This result may be explained by the reciprocal relationship between WM and PI, i.e. better resistance to PI might hinder information that is no more relevant from entering WM, and therefore in turn might enhance its capacity (May *et al.*, 1999). Conversely, higher WM capacity is associated with increased resistance to PI, as attention control seems to be more effective (Rosen and Engle, 1998; Kane and Engle, 2000; Mecklinger *et al.*, 2003). Thus, not the direct training of resistance to PI, but the higher WM performance in both tasks and both groups at the end of the training might have led to better resistance to PI, although this improvement did not spill over to other tasks that measure inhibition or resistance to interference. The outcomes of our study and the resultant assumptions should be replicated with a passive control to exclude re-test and expectancy effects. As the sample size of this study was rather small, generalization of the results may be limited and further research is needed on how mechanisms of cognitive control may underlie WM training and transfer in older adults.

The present study on WM training and PI is one among few that did not only investigate the scope of training and transfer effects, but also attempted to shed light on the cognitive processes that might underlie training and transfer. Understanding if and how WM training works is important before applying it as an intervention method to clinical populations such as neurological patients, e.g. after stroke, or to patients with mild cognitive impairment (MCI), or to psychiatric patients with WM deficits. Our results suggest that PI-demands in the training task might not be a relevant factor when developing more efficient trainings regarding WM performance and resistance to PI, but that resistance to PI can possibly be improved as a side-effect of WM training, even when it contains only low PI-demands.

Conflict of interests

None.

Description of author's roles

Study concept and design: SVL, RF, JMU, CW, BR, CPK. Development of study tasks: CPK. Data collection: SVL, RF. Data analysis: SVL, RF, CPK. Data interpretation: SVL, RF, JMU, CW, BR, CPK. Writing of the manuscript: SVL. Critical review of the manuscript: SVL, RF, JMU, CW, BR, CPK.

Acknowledgments

The authors thank Jendrik Seipp for programming of the training tasks. This research was partly supported by grants of the German Federal Ministry of Education and Research (BMBF; grant number 01GW0710), the Freiburg University Scientific Society ("Wissenschaftliche Gesellschaft Freiburg i. Br."), and the BrainLinks-BrainTools Cluster of Excellence funded by the German Research Foundation (DFG; grant number EXC 1086).

Supplementary Material

To view supplementary material for this paper, please visit http://dx.doi.org/10.1017/S1041610215001519

References

Baddeley, A. (1997). *Human Memory: Theory and Practice*. Hove: Psychology Press.

Bomyea, J. and Amir, N. (2011). The effect of an executive functioning training program on working memory capacity and intrusive thoughts. *Cognitive Therapy and Research*, 35, 529–535. doi:10.1007/s10608-011-9369-8.

Borella, E., Carretti, B., Cantarella, A., Riboldi, F., Zavagnin, M. and De Beni, R. (2014). Benefits of training visuospatial working memory in young-old and old-old. *Developmental Psychology*, 50, 714–727. doi:10.1037/a0034293.

Borella, E., Carretti, B. and Mammarella, I. C. (2006). Do working memory and susceptibility to interference predict individual differences in fluid intelligence? *European Journal of Cognitive Psychology*, 18, 51–69. doi:10.1080/09541440500215962.

Borella, E., Carretti, B., Riboldi, F. and De Beni, R. (2010). Working memory training in older adults: evidence of transfer and maintenance effects. *Psychology and Aging*, 25, 767–778. doi:10.1037/a0020683.

Bunting, M. (2006). Proactive interference and item similarity in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 183–196. doi:10.1037/0278-7393.32.2.183.

Burgess, G. C., Gray, J. R., Conway, A. R. A. and Braver, T. S. (2011). Neural mechanisms of interference control underlie the relationship between fluid intelligence and working memory span. *Journal of Experimental Psychology: General*, 140, 674–692. doi:10.1037/a0024695.

Carretti, B., Borella, E., Fostinelli, S. and Zavagnin, M. (2013a). Benefits of training working memory in amnestic mild cognitive impairment: specific and transfer effects. *International Psychogeriatrics*, 25, 617–626. doi:10.1017/S1041610212002177.

Carretti, B., Borella, E., Zavagnin, M. and de Beni, R. (2013b). Gains in language comprehension relating to working memory training in healthy older adults. *International Journal of Geriatric Psychiatry*, 28, 539–546. doi:10.1002/gps.3859.

Cohen, J. (1988). *Statistical Power Analysis for the Social Sciences*, 2nd edn. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cycowicz, Y. M., Friedman, D., Rothstein, M. and Snodgrass, J. G. (1997). Picture naming by young children: norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology*, 65, 171–237. doi:10.1006/jecp.1996.2356.

Dahlin, E., Nyberg, L., Bäckman, L. and Stigsdotter Neely, A. (2008a) Plasticity of executive functioning in young and older adults: immediate training gains, transfer, and long-term maintenance. *Psychology and Aging*, 23, 720–730. doi:10.1037/a0014296.

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

Daneman, M. and Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450–466. doi:10.1016/S0022-5371(80)90312-6.

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology*, 34, 917–928. doi:10.1093/jpepsy/jsp004.

Emery, L., Hale, S. and Myerson, J. (2008). Age differences in proactive interference, working memory, and abstract reasoning. *Psychology and Aging*, 23, 634–645. doi:10.1037/a0012577.

Gray, J. R., Chabris, C. F. and Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience*, 6, 316–322. doi:10.1038/nn1014.

Harrison, T. L., Shipstead, Z., Hicks, K. L., Hambrick, D. Z., Redick, T. S. and Engle, R. W. (2013). Working memory training may increase working memory capacity but not fluid intelligence. *Psychological Science*, 24, 2409–2419. doi:10.1177/0956797613492984.

Hartshorne, J. K. (2008). Visual working memory capacity and proactive interference. *PLoS ONE*, 3, e2716. doi:10.1371/journal.pone.0002716.

Hasher, L., Chung, C., May, C. P. and Foong, N. (2002). Age, time of testing, and proactive interference. *Canadian Journal of Experimental Psychology*, 56, 200–207. doi:10.1037/h0087397.

Hedges, L. V. and Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Orlando, FL: Academic Press.

Henson, R. N. A., Shallice, T., Josephs, O. and Dolan, R. J. (2002). Functional magnetic resonance imaging of proactive interference during spoken cued recall. *Neuroimage*, 17, 543–558. doi:10.1006/nimg.2002.1229.

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

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

Jolles, D. D. and Crone, E. A. (2012). Training the developing brain: a neurocognitive perspective. *Frontiers in Human Neuroscience*, 6, 1–13. doi:10.3398/fnhum.2012.00076.

Jonides, J., Marshuetz, C., Smith, E. E., Reuter-Lorenz, P. A., Koeppe, R. A. and Hartley, A. (2000). Age differences in behavior and PET activation reveal differences in interference resolution in verbal working memory. *Journal of Cognitive Neuroscience*, 12, 188–196. doi:10.1162/089892900561823.

Jonides, J. and Nee, D. E. (2006). Brain mechanisms of proactive interference in working memory. *Neuroscience*, 139, 181–193. doi:10.1016/j.neuroscience.2005.06.042.

Kane, M. J. and Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 336–358. doi:10.1037/0278-7393.26.2.336.

Kane, M. J. and Engle, R. W. (2003). Working memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, 132, 47–70. doi:10.1037/0096-3445.132.1.47.

Klingberg, T. *et al.* (2005). Computerized training of working memory in children with ADHD – A randomized, controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, 44, 177–186. doi:10.1097/00004583-200502000-00010.

Köstering, L., Stahl, C., Leonhart, R., Weiller, C. and Kaller, C. P. (2014). Development of planning abilities in normal aging: differential effects of specific cognitive demands. *Developmental Psychology*, 50, 293–303. doi:10.1037/a0032467.

Kyllonen, P. C. and Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, 14, 389–433. doi:10.1016/S0160-2896(05)80012-1.

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

Loosli, S. V., Rahm, B., Weiller, C., Unterrainer, J. M. and Kaller, C. P. (2014). Developmental change in proactive interference across the lifespan: evidence from two working memory tasks. *Developmental Psychology*, 50, 1060–1072. doi:10.1037/a0035231.

May, C. P., Hasher, M. and Kane, M. J. (1999). The role of interference in memory span. *Memory and Cognition*, 27, 759–767. doi:10.3758/BF03198529.

McCabe, J. and Hartman, M. (2008). Working memory for item and temporal information in younger and older adults. *Aging, Neuropsychology, and Cognition*, 15, 574–600. doi:10.1080/13825580801956217.

Mecklinger, A., Weber, K., Gunter, T. C. and Engle, R. W. (2003). Dissociable brain mechanisms for inhibitory control: effects of interference content and working memory capacity. *Cognitive Brain Research*, 18, 26–38. doi:10.1016/j.cogbrainres.2003.08.008.

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

Milham, M. P. *et al.* (2001). The relative involvement of anterior cingulate and prefrontal cortex in attentional control depends on nature of conflict. *Cognitive Brain Research*, 12, 467–473. doi:10.1016/S0926-6410(01)00076-3.

Monsell, S. (1978). Recency, immediate recognition memory, and reaction time. *Cognitive Psychology*, 10, 465–501. doi:10.1016./0010-0285(78)90008-7.

Nelson, J. K., Reuter-Lorenz, P.A., Persson, J., Sylvester, C.-Y. C. and Jonides, J. (2009). Mapping interference resolution across task domains: a shared control process in left inferior frontal gyrus. *Brain Research*, 1256, 92–100. doi:10.1016/j.brainres.2008.12.001.

Oelhafen, S., Nikolaidis, A., Padovani, T., Blaser, D., Koenig, T. and Perrig, W. J. (2013). Increased parietal activity after training of interference control. *Neuropsychologia*, 51, 2781–2790. doi:10.1016/j.neuropsychologia.2013.08.012.

Persson, J., Larsson, A. and Reuter-Lorenz, P. A. (2013). Imaging fatigue of interference control reveals the neural basis of executive resource depletion. *Journal of Cognitive Neuroscience*, 25, 338–351. doi:10.1162/jocn_a_00321.

Persson, J. and Reuter-Lorenz, P. A. (2008). Gaining control: training executive function and far transfer of the ability to the ability to resolve interference [retracted]. *Psychological Science*, 19, 881–888. doi:10.1111/j.1467-9280.2008.02172.x.

Persson, J. and Reuter-Lorenz, P. A. (2011). Retraction of "Gaining control: training executive function and far transfer of the ability to the ability to resolve interference." *Psychological Science*, 22, 562. doi:10.1177/0956797611404902.

Persson, J., Welsh, K. M., Jonides, J. and Reuter-Lorenz, P. A. (2007). Cognitive fatigue of executive processes: interaction between interference resolution tasks. *Neuropsychologia*, 45, 1571–1579. doi:10.1016/j.neuropsychologia.2006.12.007.

Postle, B. R., Brush, L. R. and Nick, A. M. (2004). Prefrontal cortex and the mediation of proactive interference in working memory. *Cognitive, Affective, and Behavioral Neuroscience*, 4, 600–608. doi:10.3758./CABN.4.4.600.

Redick, T. S. *et al.* (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.

Richmond, L. L., Morrison, A. B., Chein, J. and Olson, I. R. (2011). Working memory training and transfer in older adults. *Psychology and Aging*, 26, 813–822. doi:10.1037/a0023631.

Rosen, V. M. and Engle, R. W. (1998). Working memory capacity and suppression. *Journal of Memory and Language*, 39, 418–436. doi:10.1006/jmla.1998.2590.

Rowe, G., Hasher, L. and Turcotte, J. (2008). Age differences in visuospatial working memory. *Psychology and Aging*, 23, 79–84. doi:10.1037/0882-7974.23.1.79.

Salomon, G. and Perkins, D. N. (1989). Rocky roads to transfer: rethinking mechanisms of a neglected phenomenon. *Educational Psychologist*, 24, 113–142. doi:10.1207/s15326985ep2402-1.

Schmiedek, F., Li, S.-C. and Lindenberger, U. (2009). Interference and facilitation in spatial working memory: age-associated differences in lure effects in the n-back paradigm. *Psychology and Aging*, 24, 203–210. doi:10.1037/a0014685.

Shipstead, Z., Redick, T. S. and Engle, R. W. (2010). Does working memory training generalize? *Psychologica Belgica*, 50, 245–276. doi:10.5334/pb-50-3-4-245.

Snodgrass, J. and Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 174. doi:10.1037/0278-7393.6.2.174.

Stanislaw, H. and Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, and Computers*, 31, 137–149. doi:10.3758/BF03207704.

Studer-Luethi, B., Jaeggi, S. M., Buschkuehl, M. and Perrig, W. (2012). Influence of neuroticism and conscientiousness on working memory training outcome. *Personality and Individual Differences*, 53, 44–49. doi:10.1016/j.paid.2012.02.012.

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K. and Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proceedings of the National Academy of Sciences of the United States of America*, 94, 14792–14797. doi:10.1073/pnas.94.26.14792.

Unsworth, N. (2010). Interference control, working memory capacity, and cognitive abilities: a latent variable analysis. *Intelligence*, 38, 255–267. doi:10.1016/j.intell.2009.12.003.

Von Bastian, C. C., Langer, N., Jäncke, L. and Oberauer, K. (2013). Effects of working memory training in young and old adults. *Memory and Cognition*, 41, 611–624. doi:10.3758/s13421-012-0280-7.

Wass, S. V., Scerif, G. and Johnson, M. H. (2012). Training attentional control and working memory–Is younger, better? *Developmental Review*, 32, 360–387. doi:10.1016/j.dr.2012.07.001.

Willis, S. L. (2001). Methodological issues in behavioral intervention research with the elderly. In J. E. Birren and K. W. Schaie (eds.), *Handbook of the Psychology of Aging*, 5th edn, (pp.78–108). New York: Academic Press.

Yesavage, J. A. and Sheikh, J. I. (1986). Geriatric Depression Scale (GDS): recent evidence and development of a shorter version. *Clinical Gerontologist*, 5, 165–172. doi:10.1300/J018v05n01_09.

Zinke, K., Zeintl, M., Rose, N. S., Putzmann, J., Pydde, A. and Kliegel, M. (2014). Working memory training and transfer in older adults: effects of age, baseline performance, and training gains. *Developmental Psychology*, 50, 304–315. doi:10.1037/a0032982.