



Examining the Trainability and Transferability of Working-Memory Gating Policies

Katrina Sabah¹ · Nachshon Meiran² · Gesine Dreisbach¹

Received: 28 August 2020 / Accepted: 7 January 2021 / Published online: 27 January 2021
© The Author(s) 2021

Abstract

Internal working memory (WM) gating control policies have been suggested to constitute a critical component of task-sets that can be learned and transferred to very similar task contexts (Bhandari and Badre (*Cognition*, 172, 33–43, 2018)). Here, we attempt to expand these findings, examining whether such control policies can be also trained and transferred to other untrained cognitive control tasks, namely to task switching and AX-CPT. To this end, a context-processing WM task was used for training, allowing to manipulate either input (i.e., top-down selective entry of information into WM) or output (i.e., bottom-up selective retrieval of WM) gating control policies by employing either a context-first (CF) or context-last (CL) task structure, respectively. In this task, two contextual cues were each associated with two different stimuli. In CF condition, each trial began with a contextual cue, determining which of the two subsequent stimuli is target relevant. In contrast, in the CL condition the contextual cue appeared last, preceded by a target and non-target stimulus successively. Participants completed a task switching baseline assessment, followed by one practice and six training blocks with the WM context-processing training task. After completing training, task-switching and AX-CPT transfer blocks were administered, respectively. As hypothesized, compared to CL training condition, CF training led to improved task-switching performance. However, contrary to our predictions, training type did not influence AX-CPT performance. Taken together, the current results provide further evidence that internal control policies are (1) inherent element of task-sets, also in task switching and (2) independent of S-R mappings. However, these results need to be cautiously interpreted due to baseline differences in task-switching performance between the conditions (overall slower RTs in the CF condition). Importantly though, our results open a new venue for the realm of cognitive enhancement, pointing here for the first time to the potential of control policies training in promoting wider transfer effects.

Keywords Cognitive training · WM gating policies · Task switching · Transfer · Executive functions

Cognitive enhancement studies are faced with the challenge to produce transferable and durable learning effects to untrained structurally dissimilar contexts (i.e., far transfer) and real-life situations (Dougherty et al. 2016; Melby-Lervåg et al. 2016; Soveri et al. 2017). Despite early encouraging results, the effectiveness of cognitive training (CT) remains disputable, with current training protocols showing at best rather limited generalization effects to very similar tasks (i.e., near transfer; Karbach and Verhaeghen 2014; Schwaighofer et al. 2015).

A central pitfall of current CTs seems to be anchored in their repetitive nature, that may lead to automatization and task-specific learning, potentially incurring transfer costs (Sabah et al. 2018). In contrast, growing evidence suggests that the occurrence of learning generalization to novel contexts relies heavily on the human brain's ability to learn and represent abstract knowledge (e.g., task rules) and flexibly exploit this knowledge across multiple unfamiliar contexts, allowing rapid adjustment to new demands and situations (e.g., Badre et al. 2010; Cole et al. 2011, 2013; Collins and Frank 2013; Dreisbach 2012). Indirect evidence for the contribution of abstract task representation to learning and transfer emerges from CT studies, advocating the importance of training variability (Gopher et al. 1989; Karbach and Kray 2009; Sabah et al. 2018). For example, recent evidence suggests that the commonly undertaken approach of training “more of the same” promotes bottom-up learning, limited to

✉ Katrina Sabah
Katrina.Sabah@psychologie.uni-regensburg.de

¹ Department for Experimental Psychology, Regensburg University, Universitätstraße 31, 93053 Regensburg, Germany

² Ben Gurion University of the Negev, Beersheba, Israel

the trained task (Sabah et al. 2018). In contrast, changing task information such as task rules and stimuli throughout training (i.e., content variability) was shown to promote transfer gains and to prevent negative transfer (costs), presumably by encouraging the formation of abstract task-rules (Karbach and Kray 2009; Pereg et al. 2013; Sabah et al. 2018; Shaha et al. 2018). Additional support for these claims comes from video game training studies, attributing the favorable transfer outcomes to variability in contextual information and to the related variability in cognitive processes offered (for review see Bavelier et al. 2012).

A form of abstract knowledge that is pertinent for task execution, are *internal control policies* or *task models*—a mental program that organizes task-relevant information including rules, facts, stimuli, responses, and timing in WM to control current behavior (Bhandari and Badre 2018; Duncan et al. 2008). Here, we aim to investigate more directly how the training of internal control policies may influence transfer to new tasks that might benefit from its application.

Rules We Cannot See or Hear: The Trainability and Transferability of Working Memory Gating Control Policies

Internal control policies or task models are considered to encompass critical information for task execution, such as facts, rules, and task requirements, enabling real-time cognitive adjustments to task's dynamical structure (Bhandari and Badre 2018; Bhandari and Duncan 2014; Duncan et al. 2008). Within the domain of WM, the operation of such control policies can be embodied by gating mechanisms, which regulate the flow, updating, and maintenance of information in alignment to task dynamics through the work of input and output gates (Chatham and Badre 2015; Frank et al. 2001; Frank and Badre 2012; O'Reilly and Frank 2006; Todd et al. 2009). According to the gating framework, information flow to WM is controlled by an input gate that selects which information is to be entered and updated in WM while an output gating determines which information held in WM is response relevant. Recently, WM gating control policies were suggested to be trainable, supporting flexible behavior and learning generalization (Bhandari and Badre 2018). To train and compare between input and output gating policy training, these authors employed a second-order context processing task (see Fig. 1).

In this task, two contextual higher-order items (numbers) were each associated with two lower-order items. The cue 11 was associated with the letters A and G whereas the cue 53 was associated with the symbols \odot and π . Each trial presented a sequence of three items: a number cue, a letter, and a symbol. Which of the two items (letter or symbol) was response relevant in a given sequence was determined by the number cue. Hence, whenever the cue “11” appears, participants need

to respond to the letter. In contrast, whenever the cue “53” appears, a response to the symbol is required. To manipulate input and output gating policies, the number cue that discriminated the response-relevant item either occurred first (context-first, CF) or last (context-last, CL). As such, the early appearance of the cue enables the selective entry of the target item into WM. In the given context-first (CF) example, the first screen (“11”) indicates that the response relevant item was a letter (Fig. 1, upper panel). Response was determined in the last screen, here requiring a left-key response because the letter “A” appeared on the lower *left* side of the response panel. Conversely, in the CL condition, the number cue appeared last in the sequence, supporting the usage of output gating processes, as it requires a selective retrieval of the relevant target item (Fig. 1, lower panel). Here, the cue (53) appeared last after being preceded by the lower-order items G and \odot . Here, the response relevant item was the symbol (\odot). Because the response-relevant item (\odot) appeared on the lower *right* side of the response panel, a right-key response is required. The outstanding finding in Bhandari and Badre (2018) was that experience with either the CL or CF conditions led to transfer of the trained gating policy to new contexts with the same (e.g., CF \rightarrow CF) and different structure (e.g., CL \rightarrow CF).

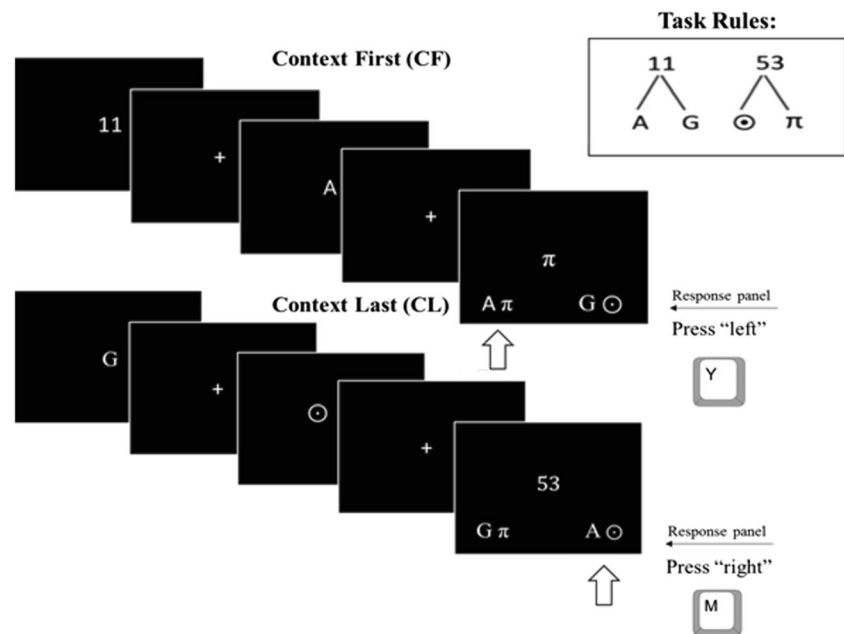
The current study attempted to examine whether a similar short-term training in WM gating policies will produce a wider transfer effect to other cognitive control tasks sharing similar task dynamics. To this end, we used Bhandari and Badre's task (2018) for training, examining possible transfer effects to cued task switching and to a context processing paradigm (the AX-continuous performance task (AX-CPT)).

Examining the Contribution of WM Gating Policies to Task-Switching and AX-CPT Performance

Task-switching (for reviews, see Kiesel et al. 2010; Monsell 2003; Vandierendonck et al. 2010) and AX-CPT (e.g., Braver and Cohen 2000; Braver et al. 2009; Hefer and Dreisbach 2016, 2017; Paxton et al. 2008) are prominent paradigms to study cognitive control processes, underlying goal-directed and flexible behavior (for reviews see Braver 2012; Gratton et al. 2018). Importantly, both the task-switching and AX-CPT paradigms were suggested to involve gating processes (Braver and Cohen 2000; D'Ardenne et al. 2012; Kessler 2017; Kessler et al. 2017; Rougier and O'Reilly 2002), making them adequate transfer tasks for the current study.

The task switching paradigm is a widely used measure for cognitive flexibility and the main measure being the latency and accuracy rates when switching as compared to repeating tasks. A popular variant is the cued task-switching paradigm, in which a cue announces which of the two tasks has to be

Fig. 1 The second-order WM task rules and structure



executed in response to a bivalent stimulus (that in principle allows the application of both tasks; Meiran 2014). Cue-processing encourages proactive control (comparable to a CF condition) and thus eases the selection of the appropriate task. The AX-CPT is a context-processing task, applied for the study of WM processes and cognitive control dynamics. In this paradigm, a target response is required whenever an A-cue is followed by an X-probe, with AX sequences occurring with high frequency (70%), making the A-cue highly predictive of the X probe. That way, the A-cue processing (comparable to a CF condition) encourages a proactive control mode leading to increased behavioral costs (higher error rates) when the A-cue is not followed by an X probe (i.e., AY trials) and to less errors when the X-probe is not preceded by an A (i.e., BX trials). Note that the typical assumption is that using a selective retrieval of contextual information is assumed to lead to higher interference on BX trials as the X-probe is strongly associated with a target response (cf. Gonthier et al. 2016).

Cued task switching and the AX-CPT thus share task dynamics with the second-order WM task, used by Bhandari and Badre (2018). Namely, they require hierarchical task representation, in which response selection is bound to a higher-order contextual cue, the task cue or the A-cue, respectively (Braver and Barch 2002). This structural similarity might facilitate transfer of cognitive control policies, here learned after exposure to the second-order WM task. For this purpose, a short-term WM gating policies training was applied, manipulating the type of trained gating policy by assigning participants to either a CF (input gating policy training) or a CL condition (output gating policy training). Task-switching performance was assessed prior and after training, utilizing a cued bivalent variation of the task-switching paradigm. Ultimately,

participants were presented with an AX-CPT transfer block. It is worth noting that we decided against an AX-CPT block prior training because the AX-CPT itself can be seen as a paradigm that promotes proactive control and cue usage like the CF condition does (for a detailed review on time on task effects in the AX-CPT, see Hefer and Dreisbach 2020). For the same reason, the order of task-switching and AX-CPT was not counterbalanced, but AX-CPT always was presented last. The following predictions were made:

1. For task-switching, we hypothesized that a selective operation of input gating control policies should encourage a selective entry of information within WM by means of enhanced cue processing, allowing for advance preparation. As such, higher transfer gains are expected to occur following CF as compared to CL training, reflected in higher overall reduction in RTs. This is supported by line of research suggesting benefits of enhanced cue processing and preparation processes to task switching performance (e.g., Meiran 1996; Savine and Braver 2010; for review see Kiesel et al. 2010). When assuming a switch-specific advanced preparation benefit, it is then expected that a higher reduction on switches when compared to repetitions will be observed, leading to a reduction in switch costs in the CF but not the CL condition (e.g., De Jong 2000; Rogers and Monsell 1995 but see Dreisbach et al. 2002; Meiran et al. 2008; Shahar and Meiran 2014; Sohn and Anderson 2001).
2. For the AX-CPT, CL training was presumed to promote selective output gating policies that promote enhanced reactive control mode, leading to lower errors rates on AY trials in the CL condition and to higher error rates

on the BX trials in comparison to the CF condition. Moreover, CF might increase usage of the A-cue, thereby leading to higher AY errors as compared to the CL condition.

Method

Participants

Eighty Regensburg University students (16 males; $M_{\text{age}} = 21.96$, $SD = 2.74$) were compensated with either 1-h course credit or were paid €6. All participants reported having normal or corrected-to-normal vision and gave written consent prior to their participation in the study.

Although the number of participants was not based on power analysis conducted in advance, we note that Bayes Factors Design Analysis (REF; Schönbrodt and Wagenmakers 2018) assuming a between-subjects effect of Cohen's $D = 0.5$ indicates the following (we used this web page for the analysis—<http://shinyapps.org/apps/BFDA/>): If the effect were present, we would have correctly detected it in 84.6% of the cases (roughly analogous to Power), would have wrongly concluded that it is absent in 0.8%, and would have remained undecided in 14.6%. If the effect were absent, we would have correctly accepted H_0 in 77% of the cases, wrongly accept H_1 (somewhat analogous to alpha errors) in 1.4%, and would have remained undecided in 21.6%.

Apparatus and Task Design

All experimental tasks were programmed in E-prime (Psychology Software Tools, Pittsburgh, PA, USA). The experiment was controlled by Dell computer with a 19" flat screen.

The Second-Order WM Control Task (Training). The task was adapted from Bhandari and Badre (2018) based on the work of Chatham and Badre (2015). On each trial, participants were presented with three stimuli, appearing in sequential order (see Fig. 1). Each sequence was composed of a number cue (11 or 53), a letter (A or G), and a symbol (π or \odot). The contextual cue (number) determined for each trial whether the letter or symbol (lower level items) was response relevant. Specifically, participants were instructed to memorize two rules through which the number 11 was associated with the letters whereas the number 53 was associated with the symbols (please see Fig. 1). For example, in a sequence composed of $11 \rightarrow G \rightarrow \odot$, the response-relevant item was the letter G whereas the symbol \odot was response irrelevant. Alternatively, in a sequence composed of $53 \rightarrow A \rightarrow \pi$, the response-relevant item was the symbol π and the letter A was response

irrelevant. Simultaneously with the presentation of the last item in a sequence, a response panel appeared on the lower part of the screen. On each side of the response panel, two pairs of lower level items, each comprising a letter and a symbol, appeared. Participants were asked to press either a left (y) or right (m) response key, depending on where the target appeared in the response panel. For example, in the sequence $11 \rightarrow G \rightarrow \odot$, "G" was the target. Hence, the required response was the (right/left) side on which "G" appeared. In the response panel, there was an equal chance of congruent and incongruent item arrangement. In a congruent arrangement, the target and irrelevant item which appeared in a sequence appeared together on the same side of the response panel, hence both associated with the same response key. For example, the arrangement in the context-first (upper) panel of Fig. 1 is congruent since the target "A" and irrelevant item " π " appeared together on the left side of the response panel, both affording a left-key response. In contrast, in an incongruent arrangement, the target and irrelevant item which appeared in a sequence were presented on opposing sides of the response panel, each associated with different response key. As such, the context-last example in Fig. 1 (lower panel) depicts an incongruent arrangement as the target " \odot " and irrelevant item "G" are associated with opponent response keys (the right and left response key, respectively). The location of the lower level items in the response panel was randomized, each appearing equally often on either the left or right side of the screen, with half of the trials requiring a right/left target response. Which of the lower item appeared first in the sequence was also randomized and balanced. All stimuli were printed in white, on a black background. We used the same stimuli set as Bhandari and Badre (2018), extracted from <https://osf.io/exyks/>.

Training consisted of six blocks, preceded by a short practice block to familiarize participants with the task and assure that the instructions were understood. Each block started with four instruction screens, followed by 48 trials. Two possible task structures were introduced, depending on the group assignment (see general procedure). For the CF group, each trial started with the cue ("11 or "53"), followed successively by two lower-level items, one response-relevant and the other response irrelevant (one letter and one symbol). In contrast, for the CL group, the cue appeared last in the sequence.

The first two items in the sequence appeared always for 300 ms whereas presentation of the last item was terminated upon response within a window of 3000 ms. A fixation cross was presented between stimulus presentation (pseudo-randomly jittered between 600 and 1600 ms) and between trials (ITI; 500). In the practice block, feedback was provided following incorrect trials, presenting the German word "Falsch" [incorrect] printed in red in the center of the screen.

Task-Switching (Baseline and Transfer Blocks). We used a modified version of the task-switching paradigm, including only mixed-tasks blocks. As stimuli, we used bivalent picture stimuli (meaning that the stimuli were relevant for the two tasks) depicting animals and objects. The size of the pictures was 1.57" × 1.18". Participants switched between two task rules. One task rule was to classify the animals/objects as fly/cannot fly (rule 1). The other rule was to classify these stimuli as living or non/living (rule 2). We used four stimuli all affording the two task-rules that were assigned to either a left response key (y) or right response key (m) on a QWERTZ keyboard, depending on the respective category. The response key assignment was counterbalanced across participants.

In both baseline and transfer, the task-switching block started with two instructional slides presenting the task rules, followed by eight practice trials and a block of 64 experimental trials. Each trial started with a fixation cross for 500 ms, followed by a cue for 650 ms. The target stimulus was then presented, remaining on screen either until a response was given or until 3500 ms had elapsed. Feedback was only presented for errors or too slow reaction times (slower than 3500 ms).

AX-CPT (Transfer). This paradigm (Servan-Schreiber 1996) is utilized as a measure of context-dependent cognitive control processes in which the cue determines the relevant response to a consecutive lower-level item (i.e., probe). A target response was required whenever the letter A appears as a cue, followed by the letter X (AX trials), occurring 70% of all trials. Three non-target trial conditions were introduced (10% each): (1) AY condition in which the "A" cue was followed by a Y-probe (Y—all letters other than X); (2) BX condition in which the "X" probe was preceded by a B-cue (B—all letters other than A) or BY condition (the cue and probe were neither the letters A nor X). The higher frequency of the AX trials results in a strong expectation for a target response following the A-cue, leading to high error rate on AY trials.

For response collection, a left response key (y) and right response key (m) on a QWERTZ keyboard were used. The assignment of the response key to target and non-target response was counterbalanced between participants. The letters were printed in 24px Calibri Light font.

The block started with three instructional slides followed by 120 experimental trials. Each trial started with a cue (300 ms), a delay of 1500 ms which was then followed by the probe (300 ms). Participants had 1300 ms to respond. A feedback was presented for too slow reaction times (slower than 1300). The trial ended with a blank screen (ITI; 1000 ms).

General Procedure

Participants were randomly assigned to one of two equal-sized training groups: (1) CF condition and (2) CL condition. They attended a 1-h experimental session, starting with task-

switching baseline block, followed by one practice block with the WM training task, six training blocks, and subsequently with task-switching and AX-CPT transfer blocks, respectively.

Results

Data Processing

Analysis focused on error rates and response times across all experimental tasks. For the training task, we followed the same data cleaning protocol as in Bhandari et al. (2018). To calculate mean RTs, erroneous trials as well as trials on which response was faster than 250 ms were discarded (6% of overall trials). For task-switching, practice trials as well as the first experimental trial were excluded from analysis. In addition, for calculating mean RTs, erroneous trials, trials following an error and trials deviating 3 SDs from the individual participant's mean in each block and trial type, were discarded (16% of overall trials). Prior to mean median RT analysis in the AX-CPT task, error trials as well as erroneous trials were excluded (5% of overall trials).

It is noteworthy that 11 participants were replaced during data collection phase following exceptionally overall high error rate on either the training task (> 30%; CF, $n = 2$; CL, $n = 3$) or testing tasks (> 50%; task switching, $n = 2$; AX-CPT, $n = 4$). For the same reason, after replacement, one participant from the CL was excluded from analysis due to exceptionally high error rate (40%), leaving only seven analyzable trials in the switch condition for that participant. Thus, data from 79 participants entered analysis eventually, utilizing error rates and response time (RT) measures.

Baseline Differences: Task-Switching

To look for potential initial differences between the training groups, 2×2 ANOVA was conducted just for the baseline block on both RTs and error rates, with trial type (repeat or switch) as a within-subjects variable and group (CF or CL) as a between-subjects variable. For the RT data, the results revealed the typical switch cost pattern, such that slower responses were obtained on task-switch ($M = 811$ ms, 95% CI [765, 857]) as compared to task-repeat trials ($M = 743$ ms, 95% CI [703, 784], $F(1, 77) = 26.75$, $p < 0.001$, $\eta_p^2 = 0.26$, $BF_{10} > 100$). Unexpectedly, the main effect group reached significance with the Bayes Factor (BF) favoring H1, $F(1, 76) = 6.25$, $p < 0.05$, $\eta_p^2 = 0.07$, $BF_{10} = 4.80$. Participants in the CF group showed overall slower RTs as compared to the CL group. In contrast, the evidence for an interaction between group and trial type provided evidence in favor of H0, $F = 1.09$, $p = 0.30$, $BF_{10} = 0.32$.

Table 1 Main effects and interaction of the Group × Block ANOVA in the context processing training task (RTs and error rates)

| | | Statistic | <i>p</i> value | Effect size (η_p^2) | BF ₁₀ |
|-------------|---------------|------------------|----------------|----------------------------|------------------|
| RTs | Group | $F(1, 77)=93.37$ | <0.001*** | 0.55 | >100 |
| | Block | $F(5, 385)=3.13$ | 0.009** | 0.04 | 1.08 |
| | Group × Block | $F(5, 385)=2.05$ | 0.07 | 0.03 | 0.32 |
| Error rates | Group | $F(1, 77)=10.04$ | 0.002** | 0.11 | 15.17 |
| | Block | $F(5, 385)=0.86$ | 0.50 | 0.01 | 0.01 |
| | Group × Block | $F(5, 385)=0.84$ | 0.52 | 0.01 | 0.04 |

p* < .01, *p* < .001

No significant effects were obtained in error data (all *F* < 2.23, all *p* > 0.13). The BFs for trial and group were indecisive tending to favor H0, BF_{10 (Trial)} = 0.49, BF_{10 (Group)} = 0.40, whereas evidence for H0 was obtained for the interaction Group × Trial, BF₁₀ = 0.28.

Training Performance—Context Processing Task

A 2 × 6 mixed-factors ANOVA was performed with training block as a within-subjects variable (1–6) and group (CF or CL) as a between-subjects variable for both RTs and error rates (see Table 1 and Fig. 2).

As seen in Table 1, the main effect for group was significant for both RTs and error data, pointing to generally faster RTs and lower error rates in the CF as compared to the CL group, thus replicating the findings of Bhandari and Badre (2018). The BF for group was aligned with the frequentist analysis, favoring H1.

The main effect for training block was significant in the RT data only, indicating to a reduction in RTs from training block 1 to block 6. The BF for block in the RT data provided only anecdotal evidence for H1 while that for the error data provided evidence for H0. The two-way interaction Group × Training Block was neither significant in RT nor error data

(see Fig. 2 and Table 2). The respective BF was indecisive for RT but provided evidence for the H0 in the error data.

Transfer Performance—Task Switching

To examine training effects on task switching performance, a 2 × 2 mixed-factors ANOVA was conducted, with pre-test versus post-test, trial type (repeat, switch) as a within-subjects variable and group (CF, CL) as a between-subjects variable (for full table of statistics and corresponding performance summary data, see Tables 3, 4, and 5).

Response Times As seen in Table 3, the main effect of block was significant, pointing to a reduction in RT from the baseline to the transfer block. The BF provided very strong evidence for H1. In addition, the main effect for trial type reached significance, denoting the typical switch costs, namely, slower RT on switch (*M* = 731 ms, 95% CI [691, 770]) as compared to repeat trials (*M* = 680 ms, 95% CI [645, 714]). This was further confirmed by Bayesian analysis, providing evidence in favor of H1. Conversely, the main effect of group did not reach significance, obtaining only anecdotal evidence for H0.

Importantly, the two-way interaction Group × Block was found to be significant. Despite the disadvantage in RT on the

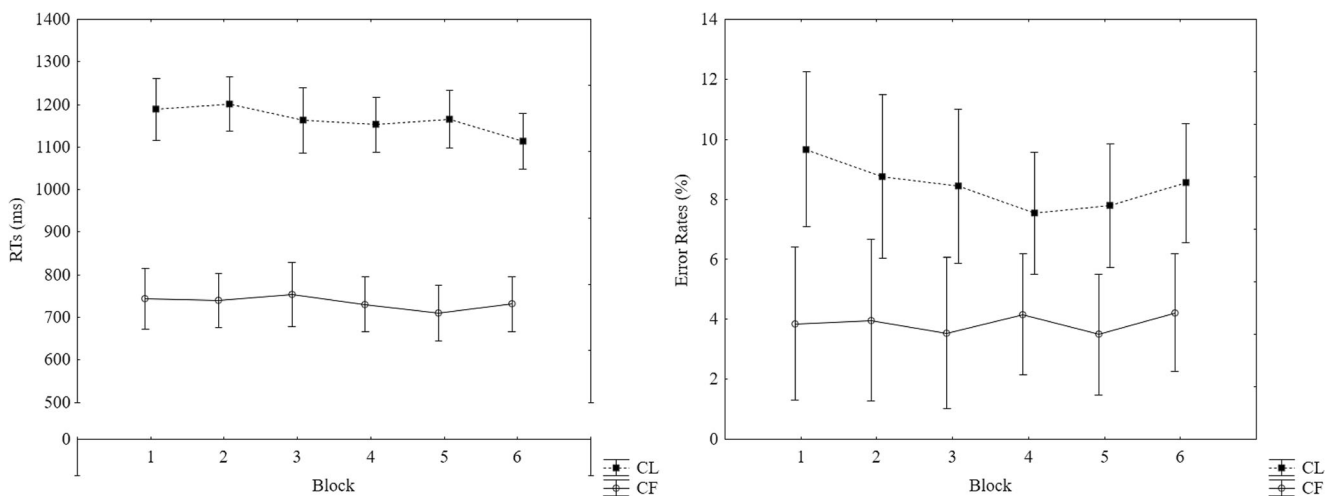


Fig. 2 Mean RT in ms (left panel) and error rates in % (right panel) across the training blocks as a function of group. Error bars denote 95% CI of the means

Table 2 Data summary for performance (accuracy, response times means and CIs) on the training task

| Group | Accuracy (errors %) | | | | | | Response times (ms) | | | | | |
|-------|---------------------|-------------|-------------|-------------|-------------|-------------|---------------------|-----------|-----------|-----------|-----------|-----------|
| | Block | | | | | | Block | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 |
| CF | 0.04 ± 0.03 | 0.04 ± 0.03 | 0.04 ± 0.03 | 0.04 ± 0.02 | 0.03 ± 0.02 | 0.04 ± 0.02 | 744 ± 71 | 740 ± 63 | 754 ± 76 | 730 ± 65 | 710 ± 66 | 731 ± 65 |
| CL | 0.10 ± 0.03 | 0.09 ± 0.03 | 0.08 ± 0.03 | 0.08 ± 0.02 | 0.08 ± 0.02 | 0.09 ± 0.02 | 1189 ± 72 | 1202 ± 64 | 1163 ± 77 | 1153 ± 65 | 1165 ± 67 | 1113 ± 66 |

baseline block as compared to the CL group, higher reduction in RTs from baseline to the transfer block was found in the CF group ($M = -193$ ms) as compared to the CL group ($M = -97$ ms), $t(77) = 2.89$, $p < 0.01$, $d = 0.65$. The BF provided very strong evidence for H1. The two-way interaction Group \times Trial Type was not significant, finding a strong evidence for H0. A significant three-way interaction was found between group trial type and block (see Fig. 3), indicating to a reduction in switch costs in the CF group, $t(39) = 3.50$, $p < 0.01$, $d = 0.55$, but not in the CL group ($p = 0.64$). In the transfer block, the CL group produced significantly higher switch costs as compared to the CF group, $t(77) = 2.07$, $p < 0.05$, $d = 0.46$. However, the BF for the three-way interaction was indecisive.

Due to pre-existing differences between the groups in overall RT, we ran an additional ANCOVA, inserting baseline RTs for repeat and switch trials as covariates, which revealed a significant two-way interaction between group and trial type, $F(1, 75) = 6.86$, $p < 0.01$, $\eta_p^2 = 0.08$.

Overall, the results indicate that CF training seems to lead to improved overall task switching performance as compared to CL training condition, indicating a transfer of learned control policies to novel untrained contexts. Moreover, our data did not provide strong support for the reduction of switch costs in the CF group as compared to the CL group.

Error Rates. The results point to significant main effect for block and very strong evidence for H1, indicating a reduction in error rates from baseline to the transfer block. Moreover, the main effect for trial type was also significant, pointing to typical switch costs, namely, higher error rates on switch compared to repeat trials. The BF for trial was indecisive, tending to favor H1. No other effect reached significance. The found BF of group and that of the interaction Trial \times Block \times Group were indecisive, favoring H0 whereas the BFs for all two-way interaction provided strong evidence for H0.

Transfer Performance—AX-CPT

A 4×2 mixed-model ANOVA was performed on both RT and error rate, including condition as a within-subject independent variable (AX, AY, BX, BY) and group (CF, CL) as a between-subject independent variable (see Tables 6 and 7 for the ANOVA results and Table 8 for performance summary data).

Response Times The results revealed the typical effect for condition, pointing to slower RT on AY as compared to AX. All other effects were not significant, with results favoring H0 (all $F < 1$, $p > 0.45$, $BF_{10} < 0.26$).

Error Rates The expected main effect for condition was significant, pointing to higher error rates on the AY as compared to

Table 3 Main effects and interaction of the Group × Block ANOVA in task-switching (RTs)

| | Statistic | <i>p</i> value | Effect size (η_p^2) | BF ₁₀ |
|----------------------------|------------------|----------------|----------------------------|------------------|
| Group | $F(1, 77)=2.93$ | 0.09 | 0.03 | 0.88 |
| Block | $F(1, 77)=77.18$ | <0.001*** | 0.50 | >100 |
| Trial Type | $F(1, 77)=29.82$ | <0.001*** | 0.28 | 24.01 |
| Group × Block | $F(1, 77)=8.34$ | 0.005** | 0.11 | >100 |
| Trial Type × Block | $F(1, 77)=7.14$ | 0.009** | 0.08 | 0.40 |
| Trial Type × Group | $F(1, 77)=0.07$ | 0.79 | 0.001 | <0.01 |
| Trial Type × Block × Group | $F(1, 77)=6.76$ | 0.01* | 0.08 | 0.56 |

p* < .05, *p* < .01, ****p* < .001

AX. The main effect of group reached significance showing higher error rates in CF group as compared to CL group. However, in contrast to our hypothesis, the two-way interaction Group × Condition did not reach significance, finding anecdotal support for H0.

Discussion

The goal of the current study was to examine whether practicing WM gating control policies can lead to beneficial transfer to different cognitive control tasks that arguably involve similar control policies. To this end, we used Bhandari and Badre’s (2018) second-order WM task and assigned participants to either an input-gating or output-gating policy training. Following training, transfer effects were assessed using the cued task-switching task and the AX-CPT task.

First, and in line with our hypothesis, CF structure training, more than CL structure training, led to improvement in task switching performance, as evidenced in RT in both switch and repeat trials. Second, we did not find any effect of CL versus CF training on performance in the AX-CPT task. There exists the possibility that this null effect reflects a methodological limitation which we were well aware of. Remember that the AX-CPT always occurred after task switching. This means that all participants had already worked through the cued task-switching blocks where they experienced a condition involving CF. As a result, whatever group differences existed beforehand could have been eliminated.

Adhering to Bhandari and Badre (2018), we choose to interpret our findings within the working-memory gating framework, attributing the advantage of the CF over CL training in task-switching to the learning and transfer of selective input-gating policies. Specifically, the early appearance of the

Table 4 Main effects and interaction of the Group × Block ANOVA in task-switching (error rates)

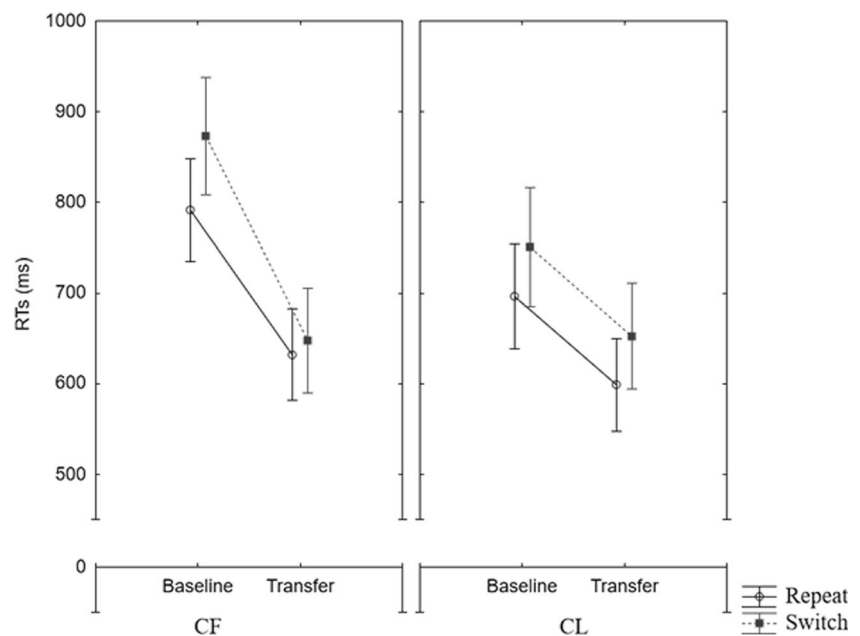
| | Statistic | <i>p</i> value | Effect size (η_p^2) | BF ₁₀ |
|----------------------------|------------------|----------------|----------------------------|------------------|
| Group | $F(1, 77)=1.51$ | 0.22 | 0.02 | 0.38 |
| Block | $F(1, 77)=52.32$ | <0.001*** | 0.40 | >100 |
| Trial Type | $F(1, 77)=8.04$ | 0.006** | 0.09 | 1.77 |
| Group × Block | $F(1, 77)=0.04$ | 0.84 | 0.001 | 0.16 |
| Trial Type × Block | $F(1, 77)=0.31$ | 0.57 | 0.004 | 0.18 |
| Trial Type × Group | $F(1, 77)=0.71$ | 0.40 | 0.01 | 0.24 |
| Trial Type × Block × Group | $F(1, 77)=0.001$ | 0.82 | 0.001 | 0.34 |

p* < .01, *p* < .001

Table 5 Data summary for performance (accuracy, response times means and CIs) on task switching

| Group | Block | Accuracy (% error) | | | Response times (ms) | | |
|-------|----------|--------------------|-------------|-------------|---------------------|----------|----------|
| | | Repeat | Switch | Overall | Repeat | Switch | Overall |
| CF | Baseline | 0.08 ± 0.03 | 0.10 ± 0.03 | 0.09 ± 0.02 | 792 ± 57 | 873 ± 65 | 833 ± 58 |
| | Transfer | 0.03 ± 0.01 | 0.05 ± 0.02 | 0.04 ± 0.01 | 632 ± 50 | 648 ± 58 | 640 ± 53 |
| CL | Baseline | 0.10 ± 0.03 | 0.11 ± 0.03 | 0.11 ± 0.02 | 696 ± 57 | 750 ± 66 | 723 ± 59 |
| | Transfer | 0.04 ± 0.01 | 0.06 ± 0.02 | 0.05 ± 0.01 | 599 ± 51 | 652 ± 59 | 626 ± 53 |

Fig. 3 Mean RT in ms in baseline and transfer task-switching blocks in the CF (left panel) and CL condition (right panel). Error bars denote 95% CI of the means



cue in the CF condition encouraged participants to exploit this contextual information in order to optimize their performance through selective selection of information into WM. Successful transfer of such form of control policy to task switching produced performance gains, presumably by promoting advance preparation for the upcoming task, shown previously to benefit task switching (e.g., De Jong et al. 1999; Dreisbach et al. 2002; Meiran 1996; Meiran and Chorev 2005; Schuch and Koch 2003). With respect to switch costs, our results were less clear. The seemingly promising and novel switch-cost reduction in the CF condition, as shown in the frequentist data analysis, was not supported by the Bayesian analysis. While such outcome can point to noise in our data, one cannot exclude the possibility that the absent Bayesian support for the alternative hypothesis is due to methodological limitation such as sample size or training dosage. Nonetheless, as argued by several authors, in the cued version of the task switching task, switch and repeat trials share preparatory processes (Dreisbach et al. 2002; Meiran et al. 2008; Shahar and Meiran 2014; Sohn and Anderson 2001). This is inherent in the unpredictable ordering of the task, requiring participants to know in advance the nature of the upcoming task (regardless if it repeats from the previous trial). These

theories thus predict a benefit of CF training in both switch and repeat trials. With regard to training dosage, it remains less clear what is the optimal training dose for effective learning and promoting transfer effects. While some advocate for the importance of high training dose for transfer emergence, others suggest to the lack of any modulating effects for training dose on transfer (Brehmer et al. 2014; Jaeggi et al. 2008; Karbach and Verhaeghen 2014; Melby-Lervåg et al. 2016; Peng and Miller 2016; Soveri et al. 2017). In fact, recent studies suggest that transfer can actually occur even after single training sessions (Sabah et al. 2018; Shahar et al. 2018). This in turn seems to be in line with existing evidence suggesting that the occurrence of skill acquisition per se requires only limited amount of practice with individuals reaching fast ceiling performance, developing automaticity (Anderson 1982; Logan 1988).

More generally, our findings provide further support that internal control policies constitute a critical component of task switching and task-sets in general, which are independent from S-R associations. This in turn bears significant implications for the study of task switching, counteracting previous theoretical models speaking against the involvement of endogenous executive control processes (e.g., task reconfiguration upon

Table 6 Main effects and interaction of the Group \times Condition ANOVA in the AX-CPT block (RTs)

| | Statistic | <i>p</i> value | Effect size (η_p^2) | BF ₁₀ |
|--------------------------|--------------------|----------------|----------------------------|------------------|
| Group | $F(1, 77)=0.65$ | 0.42 | 0.01 | 0.26 |
| Condition | $F(3, 231)=125.10$ | <0.001 *** | 0.61 | >100 |
| Group \times Condition | $F(3, 231)=0.14$ | 0.94 | 0.001 | 0.04 |

*** $p < .001$

Table 7 Main effects and interaction of the Group × Condition ANOVA in the AX-CPT block (error rates)

| | Statistic | <i>p</i> value | Effect size (η_p^2) | BF ₁₀ |
|-------------------|-------------------|----------------|----------------------------|------------------|
| Group | $F(1, 77)=8.66$ | 0.004** | 0.10 | 8.63 |
| Condition | $F(3, 231)=26.93$ | <0.001*** | 0.26 | >100 |
| Group × Condition | $F(3, 231)=2.03$ | 0.11 | 0.02 | 0.38 |

p* < .01, *p* < .001

switching and/or inhibition of previously activated task set) in cued task-switching procedures, attributing switch costs to the benefits emerging on repeat trials due to mere cue priming effects (i.e., cue repetition; Logan and Bundesen 2003; Schneider and Logan 2005). Importantly and for the first time, we were able to show that the learning of control policies is not only transferable between very similar tasks (either CL or CF tasks; Bhandari and Badre 2018) but also transfers to superficially dissimilar cognitive control tasks sharing similar task control dynamics. This in turn brings forward a new direction in cognitive training research, emphasizing the necessity to step back and reconsider the learning mechanisms of cognitive control. For example, a relevant and new theoretical contribution comes from recent studies, looking at the interaction between learning and cognitive control. This line of research points to the critical role of building up and leveraging task-set structures extracted via contextual information in the service of learning and transfer of abstract policies that support cognitive control processes (Braun et al. 2010; Collins and Frank 2013; Gershman et al. 2010; Huys et al. 2015). Interestingly, such approaches draw among others on principles of categorical learning. Specifically, in their model, Collins and Frank (2013) propose that clustering via shared similarities between higher-order contextual features, that is, how S-R contingencies are conditioned by contexts, allows to identify applicable policies across unrelated contexts. Only recently, categorization-based learning has also been claimed to play an important role in the transfer problem, enhancing the frequency of spontaneous transfer (Kurtz and Honke 2020). Others highlight the possibility that even abstract control settings such as flexibility can be learned and explained by way of associative learning (Abrahamse et al. 2016; Braem 2017; Braem and Egner 2018). As such form of learning seems to occur on higher hierarchical levels of abstraction, similar context training approaches might allow to promote wider transfer effects as

compared to the currently applied training protocols, limited to specific task-rules and S-R association (e.g., Badre et al. 2010; Bhandari and Badre 2018; Collins and Frank 2013; Frank and Badre 2012).

One possible noteworthy caveat of the current study might be the lack of control group (i.e., who did not undergo any control policy training) to decide whether it was really the CF training that improved and not the CL training that hampered task-switching performance. However, this seems implausible here when considering that in both CF and CL groups, a reduction in overall RTs from baseline to transfer was observed. An additional downside to consider is the absence of rest periods during training that might have led to fatigue, explaining as such the quite modest obtained learning effects. Instead, including rest periods has been shown to enhance cognition compared to non-rest conditions, which might have encouraged here a steeper learning curve and even stronger transfer effects (Steinborn and Huestegge 2016). An additional interesting facet to consider in future studies is the inclusion of within-subject manipulations, comparing performance on cued versus uncued trials or manipulating the preparation interval.

To conclude, our results support suggestions that internal control processes are an additional critical abstract entity constituting a task set that participants can learn through experience and reutilize across varied novel contexts (cf. Bhandari and Badre 2018). For the first time, such control policies were shown to be transferable to novel untrained cognitive control tasks (i.e., far transfer), inducing performance gains. These results open a new venue for investigation for the domain of CT, allowing to better understand possible underlying mechanisms for its effectiveness.

Funding Open Access funding enabled and organized by Projekt DEAL. Research presented in this article was funded by Deutsche Forschungsgemeinschaft (DFG), DR 392/9-1 to N.M. und G.D.

Table 8 Data summary for performance (accuracy, response times means and CIs) on the AX-CPT

| Group | Accuracy (% Error) | | | | Response times (ms) | | | |
|-------|--------------------|-------------|-------------|-------------|---------------------|----------|----------|----------|
| | AX | AY | BX | BY | AX | AY | BX | BY |
| CF | 0.22 ± 0.09 | 0.03 ± 0.08 | 0.26 ± 0.08 | 0.23 ± 0.09 | 359 ± 18 | 499 ± 21 | 345 ± 42 | 344 ± 33 |
| CL | 0.02 ± 0.09 | 0.16 ± 0.08 | 0.09 ± 0.08 | 0.05 ± 0.09 | 339 ± 18 | 490 ± 22 | 333 ± 43 | 327 ± 34 |

Data Availability The datasets generated and analysed during the current study are available in the Regensburg University publication server repository, <https://epub.uni-regensburg.de/43976/>; <https://doi.org/10.5283/epub.43976>.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from all individual participants included in the study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abrahamse, E., Braem, S., Notebaert, W., & Verguts, T. (2016). Grounding cognitive control in associative learning. *Psychological Bulletin*, *142*(7), 693–728. <https://doi.org/10.1037/bul0000047>.
- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*(4), 369–406. <https://doi.org/10.1037/0033-295X.89.4.369>.
- Badre, D., Kayser, A. S., & D'Esposito, M. (2010). Frontal cortex and the discovery of abstract action rules. *Neuron*, *66*(2), 315–326. <https://doi.org/10.1016/j.neuron.2010.03.025>.
- Bavelier, D., Green, C. S., Pouget, A., & Schrater, P. (2012). Brain plasticity through the life span: learning to learn and action video games. *Annual Review of Neuroscience*, *35*(1), 391–416. <https://doi.org/10.1146/annurev-neuro-060909-152832>.
- Bhandari, A., & Badre, D. (2018). Learning and transfer of working memory gating policies. *Cognition*, *172*, 89–100. <https://doi.org/10.1016/j.cognition.2017.12.001>.
- Bhandari, A., & Duncan, J. (2014). Goal neglect and knowledge chunking in the construction of novel behaviour. *Cognition*, *130*(1), 11–30. <https://doi.org/10.1016/j.cognition.2013.08.013>.
- Braem, S. (2017). Conditioning task switching behavior. *Cognition*, *166*, 272–276. <https://doi.org/10.1016/j.cognition.2017.05.037>.
- Braem, S., & Egner, T. (2018). Getting a grip on cognitive flexibility. *Current Directions in Psychological Science*, *27*(6), 470–476. <https://doi.org/10.1177/0963721418787475>.
- Braun, D. A., Mehring, C., & Wolpert, D. M. (2010). Structure learning in action. *Behavioural Brain Research*, *206*(2), 157–165. <https://doi.org/10.1016/j.bbr.2009.08.031>.
- Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106–113. <https://doi.org/10.1016/j.tics.2011.12.010>.
- Braver, T. S., & Barch, D. M. (2002). A theory of cognitive control, aging cognition, and neuromodulation. *Neuroscience & Biobehavioral Reviews*, *26*(7), 809–817. [https://doi.org/10.1016/S0149-7634\(02\)00067-2](https://doi.org/10.1016/S0149-7634(02)00067-2).
- Braver, T. S., & Cohen, J. D. (2000). On the control of control: the role of dopamine in regulating prefrontal function and working memory. In *Control of cognitive processes* (Vol. 18, pp. 712–737). The MIT Press. <https://doi.org/10.7551/mitpress/1481.003.0044>.
- Braver, T. S., Paxton, J. L., Locke, H. S., & Barch, D. M. (2009). Flexible neural mechanisms of cognitive control within human prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(18), 7351–7356. <https://doi.org/10.1073/pnas.0808187106>.
- Brehmer, Y., Kalpouzos, G., Wenger, E., & Lövdén, M. (2014). Plasticity of brain and cognition in older adults. *Psychological Research*, *78*(6), 790–802. <https://doi.org/10.1007/s00426-014-0587-z>.
- Chatham, C. H., & Badre, D. (2015). Multiple gates on working memory. *Current Opinion in Behavioral Sciences*, *1*, 23–31. <https://doi.org/10.1016/j.cobeha.2014.08.001>.
- Cole, M. W., Etzel, J. A., Zacks, J. M., Schneider, W., & Braver, T. S. (2011). Rapid transfer of abstract rules to novel contexts in human lateral prefrontal cortex. *Frontiers in Human Neuroscience*, *5*. <https://doi.org/10.3389/fnhum.2011.00142>.
- Cole, M. W., Laurent, P., & Stocco, A. (2013). Rapid instructed task learning: a new window into the human brain's unique capacity for flexible cognitive control. *Cognitive, Affective, & Behavioral Neuroscience*, *13*(1), 1–22. <https://doi.org/10.3758/s13415-012-0125-7>.
- Collins, A. G. E., & Frank, M. J. (2013). Cognitive control over learning: creating, clustering, and generalizing task-set structure. *Psychological Review*, *120*(1), 190–229. <https://doi.org/10.1037/a0030852>.
- D'Ardenne, K., Eshel, N., Luka, J., Lenartowicz, A., Nystrom, L. E., & Cohen, J. D. (2012). Role of prefrontal cortex and the midbrain dopamine system in working memory updating. *Proceedings of the National Academy of Sciences*, *109*(49), 19900–19909. <https://doi.org/10.1073/pnas.1116727109>.
- De Jong, R. (2000). An intention-activation account of residual switch costs. In S. Monsell & J. Driver (Eds.), *Control of Cognitive Processes* (pp. 357–376). Cambridge: MIT Press.
- De Jong, R., Berendsen, E., & Cools, R. (1999). Goal neglect and inhibitory limitations: dissociable causes of interference effects in conflict situations. *Acta Psychologica*, *101*(2–3), 379–394. [https://doi.org/10.1016/s0001-6918\(99\)00012-8](https://doi.org/10.1016/s0001-6918(99)00012-8).
- Dougherty, M. R., Hamovitz, T., & Tidwell, J. W. (2016). Reevaluating the effectiveness of n-back training on transfer through the Bayesian lens: support for the null. *Psychonomic Bulletin & Review*, *23*(1), 306–316. <https://doi.org/10.3758/s13423-015-0865-9>.
- Dreisbach, G. (2012). Mechanisms of cognitive control: the functional role of task rules. *Current Directions in Psychological Science*, *21*(4), 227–231. <https://doi.org/10.1177/0963721412449830>.
- Dreisbach, G., Haider, H., & Kluwe, R. H. (2002). Preparatory processes in the task-switching paradigm: evidence from the use of probability cues. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 468–483. <https://doi.org/10.1037/0278-7393.28.3.468>.
- Duncan, J., Parr, A., Woolgar, A., Thompson, R., Bright, P., Cox, S., Bishop, S., & Nimmo-Smith, I. (2008). Goal neglect and Spearman's g: competing parts of a complex task. *Journal of Experimental Psychology: General*, *137*(1), 131–148. <https://doi.org/10.1037/0096-3445.137.1.131>.
- Frank, M. J., & Badre, D. (2012). Mechanisms of hierarchical reinforcement learning in corticostriatal circuits 1: Computational analysis.

- Cerebral Cortex*, 22(3), 509–526. <https://doi.org/10.1093/cercor/bhr114>.
- Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001). Interactions between frontal cortex and basal ganglia in working memory: a computational model. *Cognitive, Affective, & Behavioral Neuroscience*, 1(2), 137–160. <https://doi.org/10.3758/CABN.1.2.137>.
- Gonthier, C., Macnamara, B. N., Chow, M., Conway, A. R. A., & Braver, T. S. (2016). Inducing proactive control shifts in the AX-CPT. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.01822>.
- Gopher, D., Weil, M., & Siegel, D. (1989). Practice under changing priorities: an approach to the training of complex skills. *Acta Psychologica*, 71(1–3), 147–177. [https://doi.org/10.1016/0001-6918\(89\)90007-3](https://doi.org/10.1016/0001-6918(89)90007-3).
- Gratton, G., Cooper, P., Fabiani, M., Carter, C. S., & Karayanidis, F. (2018). Dynamics of cognitive control: theoretical bases, paradigms, and a view for the future. *Psychophysiology*, 55(3), e13016. <https://doi.org/10.1111/psyp.13016>.
- Gershman, S. J., & Niv, Y. (2010). Learning latent structure: carving nature at its joints. *Current Opinion in Neurobiology*, 20(2), 251–256. <https://doi.org/10.1016/j.conb.2010.02.008>.
- Hefer, C., & Dreisbach, G. (2016). The motivational modulation of proactive control in a modified version of the AX-continuous performance task: evidence from cue-based and prime-based preparation. *Motivation Science*, 2(2), 116–134. <https://doi.org/10.1037/mot0000034>.
- Hefer, C., & Dreisbach, G. (2017). How performance-contingent reward prospect modulates cognitive control: increased cue maintenance at the cost of decreased flexibility. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(10), 1643–1658. <https://doi.org/10.1037/xlm0000397>.
- Hefer, C., & Dreisbach, G. (2020). The volatile nature of positive affect effects: opposite effects of positive affect and time on task on proactive control. *Psychological Research*, 84(3), 774–783. <https://doi.org/10.1007/s00426-018-1086-4>.
- Huys, Q. J. M., Lally, N., Faulkner, P., Eshel, N., Seifritz, E., Gershman, S. J., et al. (2015). Interplay of approximate planning strategies. *Proceedings of the National Academy of Sciences*, 112, 3098–3103.
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & 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*. <https://doi.org/10.1073/pnas.0801268105>.
- Karbach, J., & Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task-switching training. *Developmental Science*, 12(6), 978–990. <https://doi.org/10.1111/j.1467-7687.2009.00846.x>.
- Karbach, J., & Verhaeghen, P. (2014). Making working memory work: a meta-analysis of executive-control and working memory training in older adults. *Psychological Science*. <https://doi.org/10.1177/0956797614548725>.
- Kessler, Y. (2017). The role of working memory gating in task switching: a procedural version of the reference-back paradigm. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.02260>.
- Kessler, Y., Baruchin, L. J., & Bouhsira-Sabag, A. (2017). Working memory updating occurs independently of the need to maintain task-context: accounting for triggering updating in the AX-CPT paradigm. *Psychological Research*, 81(1), 191–203. <https://doi.org/10.1007/s00426-015-0717-2>.
- Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—a review. *Psychological Bulletin*, 136(5), 849–874. <https://doi.org/10.1037/a0019842>.
- Kurtz, K. J., & Honke, G. (2020). Sorting out the problem of inert knowledge: category construction to promote spontaneous transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(5), 803–821. <https://doi.org/10.1037/xlm0000750>.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*. <https://doi.org/10.1037/0033-295X.95.4.492>.
- Logan, G. D., & Bundesen, C. (2003). Clever homunculus: is there an endogenous act of control in the explicit task-cuing procedure? *Journal of Experimental Psychology: Human Perception and Performance*, 29(3), 575–599. <https://doi.org/10.1037/0096-1523.29.3.575>.
- Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6), 1423–1442. <https://doi.org/10.1037/0278-7393.22.6.1423>.
- Meiran, N. (2014). The task-cuing paradigm: a user's guide. In *Task switching and cognitive control* (pp. 45–73). Oxford University Press. <https://doi.org/10.1093/acprof:osobl/9780199921959.003.0003>.
- Meiran, N., & Chorev, Z. (2005). Phasic alertness and the residual task-switching cost. *Experimental Psychology*. <https://doi.org/10.1027/1618-3169.52.2.109>.
- Meiran, N., Kessler, Y., & Adi-Japha, E. (2008). Control by action representation and input selection (CARIS): a theoretical framework for task switching. *Psychological Research*, 72(5), 473–500. <https://doi.org/10.1007/s00426-008-0136-8>.
- Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of “far transfer”: evidence from a meta-analytic review. *Perspectives on Psychological Science*, 11(4), 512–534. <https://doi.org/10.1177/1745691616635612>.
- Monsell, S. (2003). Task switching review. *Trends in Cognitive Sciences*, 7(3), 134–140. [https://doi.org/10.1016/S1364-6613\(03\)00028-7](https://doi.org/10.1016/S1364-6613(03)00028-7).
- O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation*, 18(2), 283–328. <https://doi.org/10.1162/089976606775093909>.
- Paxton, J. L., Barch, D. M., Racine, C. A., & Braver, T. S. (2008). Cognitive control, goal maintenance, and prefrontal function in healthy aging. *Cerebral Cortex*, 18(5), 1010–1028. <https://doi.org/10.1093/cercor/bhm135>.
- Peng, P., & Miller, A. C. (2016). Does attention training work? A selective meta-analysis to explore the effects of attention training and moderators. *Learning and Individual Differences*. <https://doi.org/10.1016/j.lindif.2015.11.012>.
- Pereg, M., Shahar, N., & Meiran, N. (2013). Task switching training effects are mediated by working-memory management. *Intelligence*, 41(5), 467–478. <https://doi.org/10.1016/j.intell.2013.06.009>.
- Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, 124(2), 207–231. <https://doi.org/10.1037/0096-3445.124.2.207>.
- Rougier, N. P., & O'Reilly, R. C. (2002). Learning representations in a gated prefrontal cortex model of dynamic task switching. *Cognitive Science*, 26(4), 503–520. https://doi.org/10.1207/s15516709cog2604_4.
- Sabah, K., Dolc, T., Meiran, N., & Dreisbach, G. (2018). When less is more: costs and benefits of varied vs. fixed content and structure in short-term task switching training. *Psychological Research*, 1–12. <https://doi.org/10.1007/s00426-018-1006-7>.
- Savine, A. C., & Braver, T. S. (2010). Motivated cognitive control: reward incentives modulate preparatory neural activity during task-switching. *Journal of Neuroscience*, 30(31), 10294–10305. <https://doi.org/10.1523/JNEUROSCI.2052-10.2010>.
- Schneider, D. W., & Logan, G. D. (2005). Modeling task switching without switching tasks: a short-term priming account of explicitly cued performance. *Journal of Experimental Psychology: General*, 134(3), 343–367. <https://doi.org/10.1037/0096-3445.134.3.343>.
- Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: planning for compelling evidence. *Psychonomic Bulletin & Review*, 25(1), 128–142. <https://doi.org/10.3758/s13423-017-1230-y>.

- Schuch, S., & Koch, I. (2003). The role of response selection for inhibition of task sets in task shifting. *Journal of Experimental Psychology: Human Perception and Performance*, 29(1), 92–105. <https://doi.org/10.1037/0096-1523.29.1.92>.
- Schwaighofer, M., Fischer, F., & Bühner, M. (2015). Does working memory training transfer? A meta-analysis including training conditions as moderators. *Educational Psychologist*, 50(2), 138–166. <https://doi.org/10.1080/00461520.2015.1036274>.
- Servan-Schreiber, D. (1996). Schizophrenic deficits in the processing of context. *Archives of General Psychiatry*, 53(12), 1105.
- Shahar, N., & Meiran, N. (2014). Differential contribution of task conflicts to task switch cost and task mixing cost in alternating runs and cued task-switching: evidence from ex-Gaussian modeling of reaction time distributions. *Psychological Research*, 79(2), 259–266. <https://doi.org/10.1007/s00426-014-0569-1>.
- Shahar, N., Pereg, M., Teodorescu, A. R., Moran, R., Karmon-Presser, A., & Meiran, N. (2018). Formation of abstract task representations: exploring dosage and mechanisms of working memory training effects. *Cognition*, 181, 151–159. <https://doi.org/10.1016/j.cognition.2018.08.007>.
- Sohn, M. H., & Anderson, J. R. (2001). Task preparation and task repetition: two-component model of task switching. *Journal of Experimental Psychology: General*, 130(4), 764–778. <https://doi.org/10.1037/0096-3445.130.4.764>.
- 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 and Review*, 24(4), 1077–1096. <https://doi.org/10.3758/s13423-016-1217-0>.
- Steinborn, M. B., & Huestegge, L. (2016). A walk down the lane gives wings to your brain. Restorative benefits of rest breaks on cognition and self-control. *Applied Cognitive Psychology*. <https://doi.org/10.1002/acp.3255>.
- Todd, M. T., Niv, Y., & Cohen, J. D. (2009). Learning to use working memory in partially observable environments through dopaminergic reinforcement. In *Advances in Neural Information Processing Systems 21 - Proceedings of the 2008 Conference* (pp. 1689–1696).
- Vandierendonck, A., Liefoghe, B., & Verbruggen, F. (2010). Task switching: interplay of reconfiguration and interference control. *Psychological Bulletin*, 136(4), 601–626. <https://doi.org/10.1037/a0019791>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.