GENERALIZABILITY AND MECHANISMS OF LEARNED FLEXIBILITY INDUCED THROUGH SWITCH PROBABILITY MANIPULATION

by

Corey A. Nack

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science

Department of Psychological Sciences

West Lafayette, Indiana

May 2022

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Yu-Chin Chiu, Chair

Department of Psychological Sciences

Dr. Darryl W. Schneider

Department of Psychological Sciences

Dr. Sebastien Hélie

Department of Psychological Sciences

Approved by:

Dr. Kimberly P. Kinzig

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The brain dynamically alters its production of flexible behavior: cognitive flexibility increases when demand is high. In task switching experiments, past exposure to a high demand for flexibility in conjunction with specific temporal contexts leads to learned switch readiness such that future exposures to those contexts will cue flexibility. According to a recent proposal (Dreisbach & Fröber, 2019), learned switch readiness following switch demands is supported by a concurrent activation (CA) cognitive mechanism whereby both sets of task rules are kept available in working memory despite only using one at a time. This can be differentiated from a competing candidate mechanism, working memory updating (WMU) thresholds which determine the ease of replacing one task's rules with another. The WMU mechanism is expected to cause a global increase in flexibility while CA is conceptualized as limited to task-specific associations. To test whether learned switch readiness represents a global or limited change in the cognitive system, I conducted two experiments that both involved learning switch readiness in one context and generalizing it in another. In Experiment 1, I replicated and extended findings that switch probability manipulations can modulate voluntary switch rates (VSR), indicating one type of generalizability. However, in Experiment 2, I found that flexibility learned through switch probability manipulations did not transfer to new tasks when the task rules were changed but contextual cues remained the same, demonstrating a limit: learned switch readiness does not generalize across tasks. These findings together suggest that CA is likely the mechanism behind learned switch readiness.

INTRODUCTION

Adapting to changing demands is an important feature of cognitive systems. Driving, for example, requires people to switch between different tasks regularly. From determining whether they are driving within the posted speed limit to assessing how safe it is to move lanes, the goal of the driver requires these tasks to be performed in a regular and alternating fashion. While driving familiar paths may require less switching than normal, other more complicated routes can require more. For instance, consider an intersection where pedestrians frequently cut across roads out of turn. Having an efficient cognitive system that adjusts for high switch demands in this new environment would be advantageous. An even better system would track the demand for switching over time and learn to associate contextual cues leading up to that intersection with the increased switch demand.

This is indeed a feature of human cognition: our minds are flexible and adaptive to various cognitive demands as a result of utilizing environmental cues associated with a particular demand (e.g., Bugg & Crump, 2012; Chiu & Egner, 2017; Egner, 2014; Leboe et al., 2008). Just as learned simple stimulus-response associations can influence overt behavior, so do learned associations between a stimulus and the need for cognitive control (Abrahamse et al., 2016; Braem & Egner, 2018). This has been referred to as context-control learning (e.g., Chiu & Egner, 2019). However, the degree to which such learning will generalize beyond the learning context is unknown. The present study assesses whether learned flexibility is limited to the same context in which it was learned, or whether it can be generalized to dissimilar contexts. I begin to address this question by identifying the extent to which learned flexibility can generalize a) from a learning context where flexibility is externally cued to one wherein flexibility is a voluntary choice, and b) from a learning context where two tasks are used to teach flexibility to one wherein different tasks (with no task-specific learned flexibility) are used but the old learned flexibility cues remain. Below, I review the relevant literature and provide a rationale for the current experiments.

Switch Costs and Learned Switch Readiness

When given two tasks to perform, Task A and Task B, a person completing Task A twice in a row will have better performance than someone completing Task A and then Task B (Allport et al., 1994; Jersild, 1927; Meiran, 1996; Monsell, 2003; Rogers & Monsell, 1995; Schneider & Logan, 2009; Spector & Biederman, 1976; Wylie & Allport, 2000). The resulting difference in response time (RT) or error rates (ER) is called a switch cost. Switch costs can be decreased when switches are frequent (Dreisbach & Haider, 2006; Monsell & Mizon, 2006; Schneider & Logan, 2006). For example, switch costs can be modulated by so-called "list-wide switch probability" manipulations. In the list-wide switch probability manipulation, a frequent switch list consists of more task switch trials and fewer task repeat trials than a rare switch list with a frequency-reversed trial composition. Typically, the list with more switches produces smaller switch costs (Dreisbach & Haider, 2006; Monsell & Mizon, 2006; Schneider & Logan, 2006). This difference in switch costs between two lists with differentially biased switch probabilities is referred to as a list-wide switch probability (LWSP) effect. The interpretation of this LWSP effect is that participants adaptively increase switch readiness in a preparatory fashion across all trials in the same block as a result of encountering more switch trials that require task-rule updating during a frequent switch list, (Braver et al., 2007; Bugg & Crump, 2012).

The LWSP effect demonstrates that participants can associate particular lists with a higher demand for switching, and subsequently retrieve these associations to facilitate performance, leading to reduced switch costs. This can be described as 'learned switch readiness' (Chiu & Egner, 2017). Switch costs therefore serve as one index of learned switch readiness, as changes in costs likely indicate changes in the speeds at which the appropriate control-state is directly retrieved as a result of repeated pairings between the appropriate control-state and a particular temporal context (i.e., a list of trials). Another index of switch readiness is voluntary switch rate, or VSR, measured as the proportion of free-choice trials wherein participants chose to switch tasks (Arrington & Logan, 2004; 2005). VSR has been used to independently validate switch costs as a measure of switch readiness (Brosowsky & Egner, 2021; Chiu et al., 2020). At the level of computational modeling, switch readiness can be instantiated within models of control learning or reinforcement learning as a control parameter that incrementally changes to adapt to increases in the demand for flexibility (e.g., Botvinick et al., 2001; Jiang et al., 2014, 2015).

Regardless of how they are indexed or computationally modeled, findings like the LWSP effect indicate a major role of associative learning and bottom-up priming in guiding cognitive control at a given moment (Braem & Egner, 2018). Such evidence has led researchers to reevaluate the traditional views of cognitive control adjustments which primarily emphasize cognitive control as a top-down or volitional adjustment of responding once it becomes apparent that the habitual strategy or automatic response is disadvantageous (Diamond, 2013). In contrast to this traditional automatic/controlled dichotomy, the associative learning account of cognitive control holds that control adjustments, just like ‘lower-level’ stimulus-response associations, emerge from automatic processing of bottom-up cues (Abrahamse et al., 2016; Braem & Egner, 2018).

More generally, the degree to which a person is exhibiting flexible or stable behavior is referred to as a metacontrol state, meaning that cognitive control over behavior is itself being controlled (Goschke, 2013; Hommel, 2015). For example, the differences in the learned switch readiness between lists shown by LWSP manipulations can be described as a bottom-up shift along a metacontrol continuum towards a more flexible state. Similar to list-based switch probability cues, the identity of the stimuli themselves (Chiu & Egner, 2017; Leboe et al., 2008) or features of the stimuli like spatial location on the display (Bugg & Crump, 2012; Crump & Logan, 2010; Leboe et al., 2008) can also serve as cues to shift along the metacontrol continuum. Likewise, factors unrelated to switch probabilities, such as reward (Chiew & Braver, 2014; Fröber et al., 2019; Hefer & Dreisbach, 2016; Locke & Braver, 2008), and affective state (Dreisbach, 2006; Dreisbach & Goschke, 2004; Isen, 2001; van Wouwe et al., 2011) have been shown to shift the metacontrol state along the flexibility-stability continuum. These latter determiners of metacontrol state likely utilize a different cognitive mechanism from learned switch readiness (Dreisbach & Fröber, 2019) but this has yet to be empirically demonstrated. One means of clarifying whether the behavioral effects of these metacontrol shifts (i.e., whether induced by switch probability cues, reward, or affect) all arise from the same cognitive mechanism would be to determine whether the consequences of each metacontrol shifts represent a local or global reorganization of the cognitive system. Two modes or mechanisms of flexibility are of particular relevance to this generalizability question: proactive metacontrol and concurrent activation (CA).

Proactive and Reactive Metacontrol

According to the dual mechanisms of cognitive control framework (Braver, 2012; Braver et al., 2007), there are two distinct modes of cognitive control. Proactive control keeps top-down goal representations active throughout a task, even in the absence of immediate control demands like response conflict. Reactive control, by contrast, entails only reinstituting goal representations once control demands are encountered. Both types of control have been found to be relevant to task switching (Braem et al., 2019). While proactive and reactive control can guide behavior on a given task, these control states themselves can become associated with contextual cues and be retrieved when needed. These retrieval strategies have been referred to as proactive or reactive metacontrol (distinct from flexibility vs. stability metacontrol). LWSP effects reflect proactive metacontrol, as the participant is using the temporal cue provided by the list to retrieve the proactive control mode for a sustained period of time. This can be differentiated from reactive metacontrol, where the appropriate control mode is only activated briefly after a cue such as the identity of the stimulus itself predicts a likely control demand. The two metacontrol states have been shown to be uncorrelated, indicating that their underlying mechanism might not be the same (Kang & Chiu, 2021). A remaining question is whether proactive metacontrol can influence voluntary switch decisions in the same way it has been shown to influence cued task switching performance. This is primarily a basic science question but could also be applicable to clinical populations characterized by deficits in flexibility (e.g., Geurts et al., 2009; Meiran et al., 2011). One goal of treatment is to encourage patients to generalize from flexibility that is cued by a therapist or training program to voluntary flexibility in the absence of the therapist. Unfortunately, such flexibility interventions often show very limited impact beyond the training context (Simons et al., 2016). To the extent that generalizability from cued to voluntary flexibility has been shown, it has so far only been demonstrated for reactive metacontrol. A better understanding of the transferability of proactive metacontrol could illuminate better strategies for developing therapeutic interventions.

Concurrent Activation and Working Memory Updating in Metacontrol

Demonstrating the influence of proactive metacontrol on both cued and voluntary task switching would indicate diverse temporal mechanics in learned switch readiness (i.e., a long-term component to compliment the previously established short-term component). However, such findings would be insufficient to describe exactly how learned switch readiness is actually applied to improve switching performance. According to a recent theory proposed by Dreisbach and Fröber (2019), two separate cognitive mechanisms might contribute to such improvements. The first is a reduction in working memory updating (WMU) thresholds, meaning that it becomes easier to replace one task representation with another in working memory. The other mechanism, concurrent activation (CA), occurs when mental representations of two tasks remain active simultaneously, such that both are highly accessible at any time, regardless of which one is currently required. The authors also hypothesize that reward and affect are more likely to modulate WMU-based flexibility (e.g., Isen, 2001) whereas cues from the context of the task are more likely to modulate CA-based flexibility. The reason for this is that positive affect or unexpected rewards should signal increased exploratory behavior in general, allowing one to discover and adopt unanticipated and previously overlooked mental representations under favorable circumstances. Contextual cues of high switch demand, by contrast, should not signal a need for general exploratory behavior, but only indicate the need for both sets of task rules in rapid succession. In other words, the cognitive system responds to high demands for switching by keeping a task representation active, even when switching away from that task, based on the likelihood that the task will be performed again soon. Because of this, CA-based flexibility is expected to be task-specific as concurrent activation of two representations grants no benefit to, and could conceivably even impede, the activation of additional representations. If this is the primary mechanism driving item-specific effects, then a stimulus that invokes concurrent activation of two practiced tasks would have no switching advantage when two new tasks are performed on that stimulus. I hypothesize that CA is the primary driver of the cognitive flexibility guided by bottom-up cues (e.g., lists, items, etc.) that are associated with particular switch probabilities. As a result, it is likely that switch readiness driven by CA-based flexibility is task-specific and has limited generalizability. Therefore, demarcating the generalizability of metacontrol shifts due to context-switch probability associations (i.e., LWSP manipulations) would offer insight into the cognitive mechanisms underlying flexibility, allowing for a test of the CA mechanism in Dreisbach and Fröber's (2019) theory.

Here I present two experiments aimed at inducing switch readiness and investigating the degree to which the learned switch readiness can be retrieved outside of the learning context. In Experiment 1, I examined whether temporal switch cues modulated the choice to perform task switches, similar to how item-based cues have been shown to influence such choices. In Experiment 2, I investigated whether switch probability effects are task-specific, or if they hold when new tasks are introduced. With these experiments, I demarcated the generalizability of learned flexibility by determining whether VSR is sensitive to proactive metacontrol (Experiment 1) and by ruling out working-memory updating thresholds as the mechanism underlying cognitive flexibility guided by context-specific switch probabilities, indicating CA as the mechanism instead (Experient 2).

Experiment 1

Recently, Chiu et al. (2020) investigated learned switch readiness using an item-specific switch probability (ISSP) paradigm, the reactive-control counterpart to LWSP wherein particular stimuli, not lists of stimuli, are associated with a switch probability bias. The authors found that such control learning is not limited to modulating the switch costs in cued task switching contexts, but also generalizes to a voluntary task switching context where participants are given the opportunity to freely decide which task to perform (thus when to switch or repeat as well). This study featured cued trials for most of the experiment. However, for 25% of the trials, participants were allowed to choose which of the two tasks they wished to perform. These voluntary trials afford a second way to assess switch readiness by measuring the VSR. Results show that stimuli which frequently appeared on cued task switch trials not only had reduced switch costs on cued trials, but also had a higher VSR. This study demonstrates that under circumstances requiring frequent switching, participants can learn to become both better at and more willing to perform task switches. While it has been shown that VSR can be modulated by item-specific contextual cues, the extent to which a temporal, list-based cue can modulate VSR is unclear.

This question mostly concerns the different types of metacontrol that can be induced by the switch probability manipulations. On the one hand, the ISSP effect has been interpreted as reactive metacontrol. This is because the appropriate control state can only be triggered after the stimulus onset as the items predicting frequent or rare switching are fully intermixed and occur randomly across trials. In other words, participants cannot predict whether an upcoming trial will require high or a low switch readiness. On the other hand, the LWSP effect has been interpreted as proactive metacontrol because the appropriate control state can be prepared prior to the stimulus onset and is thought to be applied to all trials in the same list. However, to claim that the effect is truly ‘proactive,’ an experiment must include ‘diagnostic’ items with unbiased switch frequencies to demonstrate that the control state is applied to those diagnostic items, even though these items are associated with a 50% switch probability. The basic LWSP effect on switch cost has indeed been successfully replicated in diagnostic items (Kang & Chiu, 2021). Here, I will extend these efforts and examine VSR in both switch frequency-biased inducer items and frequency-unbiased diagnostic items. Such a study would demonstrate that list-wide switch probability influences switch readiness in ways beyond reducing switch costs. I hypothesized that proactive metacontrol would be induced by the list-wide switch probability manipulation. As a consequence, both switch costs and VSR should be modulated in both the biased and unbiased items. Finding a pattern of results whereby VSR is higher in a frequent switch list than a low switch list, especially if such an effect was present among diagnostic items (as has yet to be demonstrated), would support my hypothesis.

Experiment 2

In Experiment 2, I attempted to push the generalizability further and investigate whether learned flexibility is task-specific or able to transfer to new switching tasks. It is possible that item-based cuing consists of a learned association between a stimulus and a high demand for switches, regardless of the task being switched to. Some evidence of this is provided by the fact that ISSP effects hold when there are three tasks in a switching paradigm rather than the typical two (Chiu & Egner, 2017). In and of itself, however, this does not indicate task-nonspecific learning, as all three tasks were subject to equal amounts of switch probability learning. Other evidence suggests that learned switch readiness is task specific. For example, Sabah et al. (2019) conducted a training experiment wherein participants learned to switch between different tasks, and then were given two novel tasks to perform. Task performance decreased and switch costs increased when the tasks were replaced. However, this could show task specificity for learned switch readiness, or both changes could have been caused by a simple lack of practice with the new tasks. Moreover, switch probability was not manipulated in this experiment. A direct test would be informative, as Dreisbach and Fröber (2019) propose that changing tasks would result in the loss of learned switch readiness. An experiment which probes the persistence of learned flexibility once tasks are replaced with new ones would help clarify the generalizability of this effect and help test Dreisbach and Fröber's (2019) description of the underlying mechanism.

I conducted such a test by creating an LWSP effect via switch probability manipulation and then replacing the tasks with two new tasks that had not been associated with biased switch probabilities. However, the stimuli and lists used in this second 'transfer phase' remained unchanged, allowing for the associations between lists and learned switch readiness to persist. Because the switch probabilities for both lists reverted to 50% in the transfer phase, any differences in switch costs between formerly frequent switch and rare switch lists must come from associations learned before the tasks were replaced (i.e., among a *different* pair of tasks). Such task-nonspecificity would be incompatible with concurrent activation. Therefore, I expected a pattern of results whereby an LWSP effect (the interaction between switch costs and list-wide switch probability) would be present in the learning context, but not in a dissimilar transfer context with different switching tasks.

EXPERIMENT 1

Method

Participants

112 total participants (61 female, 51 male; mean age = 18.61, SD = 1.12) were recruited for a study approved by the Purdue University Institutional Review Board (IRB) in exchange for credits towards a class requirement. Of these, 17 were excluded (9 due to poor performance and 8 due to having at least one condition with no valid trials after excluding outlier trials). The following analyses and report therefore have a final sample of 95, which exceeded the preregistered sample of 80 due to having fewer participants excluded than anticipated. The target sample size was determined to provide over 95% power to detect LWSP effects (alpha = .05) based on a power simulation that sampled values for each within-subjects condition from distributions that have the same means and standard deviations reported by Chiu et al. (2020; Experiment 2). Participants were excluded if their overall error rate was more than 2.5 standard deviations from the group mean or if at least one experimental condition was found to have no valid responses. These exclusion criteria were determined ahead of data collection and preregistered at aspredicted.org (https://aspredicted.org/4MW_L5N).

Apparatus and Stimuli

This experiment was conducted in-person. The stimuli were presented on a 22-inch Dell monitor with a resolution of 1920 x 1080, controlled by JavaScript interpreted by Google Chrome. Stimuli consisted of 16 images featuring a central object on a white background (Moreno-Martínez & Montoro, 2012; Possidónio et al., 2019). Stimuli could be classified according to two dimensions: animacy (living vs. nonliving) or size (larger vs. smaller than a shoebox). Thus, 4 fully orthogonal stimuli categories were possible (living-large, living-small, nonliving-large, nonliving-small) and an equal number of items from each category were included in the experiment. An additional set of 8 images was used for practice exclusively. Stimuli were presented at 9.00 cm wide x 6.70 cm tall and were surrounded by a .30 cm thick rectangular frame that could be either red, blue, or gray.

Design and Procedure

Participants switched between an animacy categorization task (living vs. nonliving) and a size categorization task (larger vs. smaller than a shoebox). The experiment had two phases: a 100% cued phase consisting of only cued trials (i.e., participants were told which task to perform, and consequently, whether to switch or repeat) and a hybrid phase consisting of 75% cued trials and 25% free choice trials (i.e., participants freely chose which task to perform). On cued trials, a blue or a red frame around the stimulus indicated which task to perform. This frame appeared simultaneously with the stimulus, preventing advanced task preparation. On free choice trials, a gray frame appeared, and participants performed a task of their choosing.

See Figure 1 for an overview of all experiment components. The 100% cued phase consisted of two lists of 240 cued task switching trials, with one list associated with a high list-wide switch probability (65%, 'frequent switch list') and the other one associated with a low list-wide switch probability (35%, 'rare switch list'). Specifically, each list consisted of 8 unique items (two from each of the four image categories: living-large, living-small, nonliving-large, nonliving-small). Half of the stimuli were 'inducer' items and were associated with either an 80% or 20% chance of switching. In contrast, the other half of the stimuli were 'diagnostic' items and were associated with a 50% chance of switching. The strong bias in switch probability among inducer items thereby created the overall difference in switch probability between the frequent and rare switch lists. All items were unique to each list; no items from one list appeared in the other. The order of the frequent/rare lists was counterbalanced across participants. The hybrid phase consisted of two lists of 320 trials each, with each list consisting of 240 cued trials (structured exactly like the 100% cued phase) and 80 free choice trials. Placement of free choice trials was constrained so that they never appeared consecutively. If participants showed a preference for one task over the other on over 70% of free choice trials within a block, they were presented the following message at the end of the block: "Try to choose the two tasks equally often, like flipping a coin".

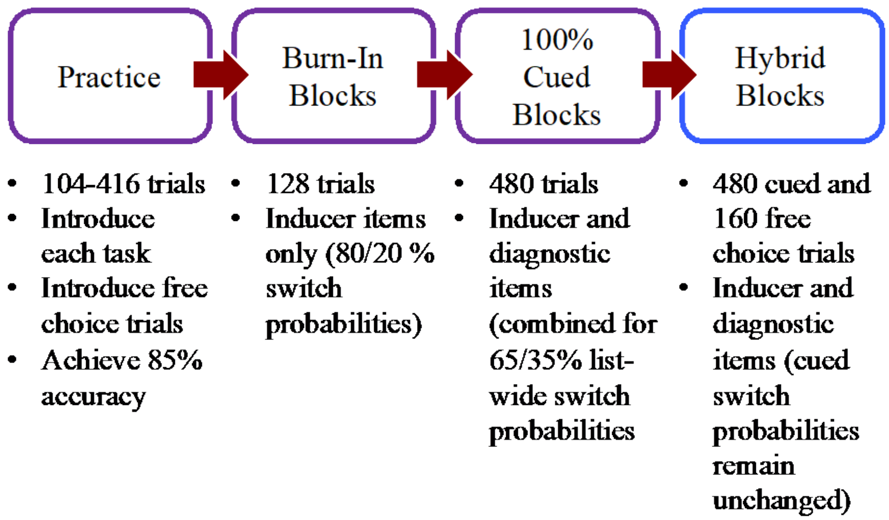


Figure 1. Experiment 1 design.

All trials began with a fixation cross for 300 ms, followed by the stimulus for 1500 ms. The deadline was increased from the proposed 1200 ms but kept relatively short to ensure the task was difficult enough to require cognitive control and to make response time distributions less skewed. Based on uncleaned response times from a previous study (Nack, under review), this deadline was deemed sufficient. All trials ended with either “Correct!,” “incorrect,” “too slow,” or “wrong key” presented as feedback for 600 ms. See Figure 2 for example trials. Participants used the A and Z keys on a QWERTY keyboard to respond to one task, and the K and M keys to respond to the other task. Note that this meant that there was a “left hand” task and a “right hand” task, allowing me to identify the intended task on free choice trials. All response key mappings and frame color-to-task-mappings were counterbalanced across participants.

Before the main experiment, participants provided informed consent and had questions answered. Next, they began with a practice phase of between 104 and 416 trials, all with no bias in switch probability. Practice continued until the maximum trial count or participants achieving a mean accuracy of at least 85%. This practice phase taught participants the response key mappings for one task, then the other, then both tasks mixed, and finally voluntary choice trials. After finishing the practice trials, participants performed two ‘burn-in blocks’ (64 trials each, one block from each list, list order counterbalanced across participants) featuring only inducer items. These burn-in blocks were not included in analysis. In total, participants completed 1208 experimental trials and were given breaks after each block to reduce fatigue.

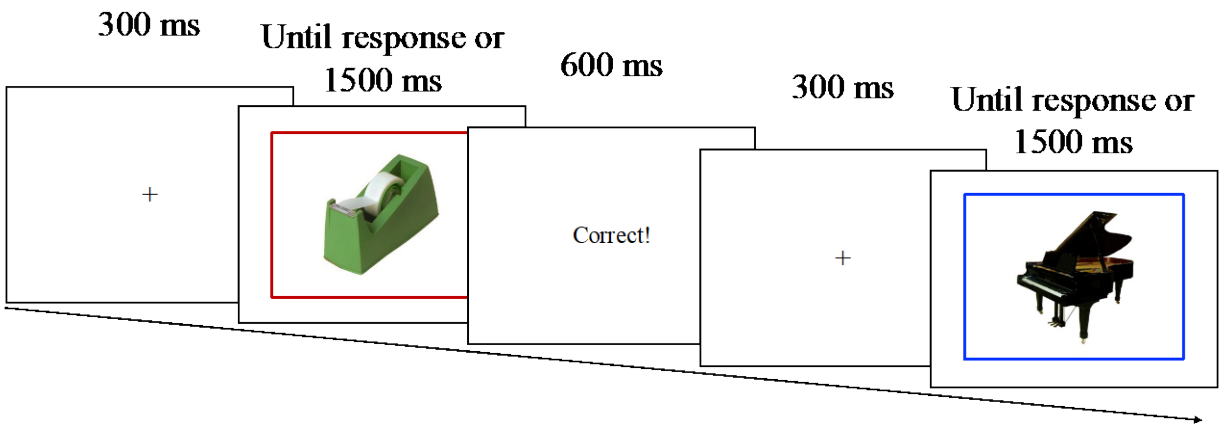


Figure 2. Exemplar trial progression.

Analysis

Prior to analysis, outlier trials were excluded. Individual trials with a response time either more than 3 SD from a participant's mean or less than 200 ms, as well the first trials of each block were excluded from all analyses. For RT and VSR analyses, error trials, trials immediately following errors and trials with no response and were excluded. For the ER analysis, trials with no response were excluded.

To address the main question of whether a list wide switch probability manipulation would modulate participants' preference for repeating over switching, VSR was subjected to a 2 (item type: inducer, diagnostic) x 2 (list: frequent switch, rare switch) repeated measures analysis of variance (ANOVA). Four additional ANOVAs investigated the list wide switch probability manipulation on RT and ER, separately for the 100% cued and hybrid phases. Each of these ANOVAs used a 2 (item type: inducer, diagnostic) x 2 (list: frequent switch, rare switch) x 2 (transition: switch, repeat) repeated measures design. Voluntary choice trials were not included in RT or ER analysis for the hybrid phase. In the event of a significant 3-way interaction that involved a list x transition interaction (the LWSP effect), we followed up with post-hoc analyses that were corrected for multiple comparisons using the Bonferroni correction. The critical goal of these follow-up investigations was to detect LWSP effects among diagnostic items.

Given that a lack of difference in the LWSP effect between inducer and diagnostic items (i.e., a nonsignificant 3-way interaction) would also support the presence of proactive metacontrol, we conducted JZS Bayes factor ANOVAs to compliment the original ANOVAs' inability to quantify evidence for invariance between conditions (Rouder et al., 2009). We reported results for the relevant interactions involving item types. Using default priors on version 0.9.12-4.2 of the BayesFactor R package (Rouder et al., 2012), I computed a BF_{01} (H0/H1) to quantify the amount of evidence for a lack of an interaction effect. Specifically, the null model (H0) for which the BF_{01} reported here quantifies evidence is the full model minus the interaction in question, while the alternative model (H1) is the full model. A Bayes factor between 1 and 3 indicates anecdotal evidence, a value between 3 and 10 indicates substantial evidence, and a value over 10 indicates strong evidence for the null model, i.e., no difference between inducer and diagnostic items (Jeffreys, 1961; Robert et al., 2009). In contrast, if the BF_{01} falls below 1, it ceases quantifying evidence for the null and instead begins to indicate support for the alternative model. We also report estimated proportional error of the Bayes factor, indicating the degree of error among Monte Carlo samples as a percentage of the associated BF value. For example, for a BF_{01} of 20 and an error ± 5%, the estimated error covers a range from 19 to 21.

Results

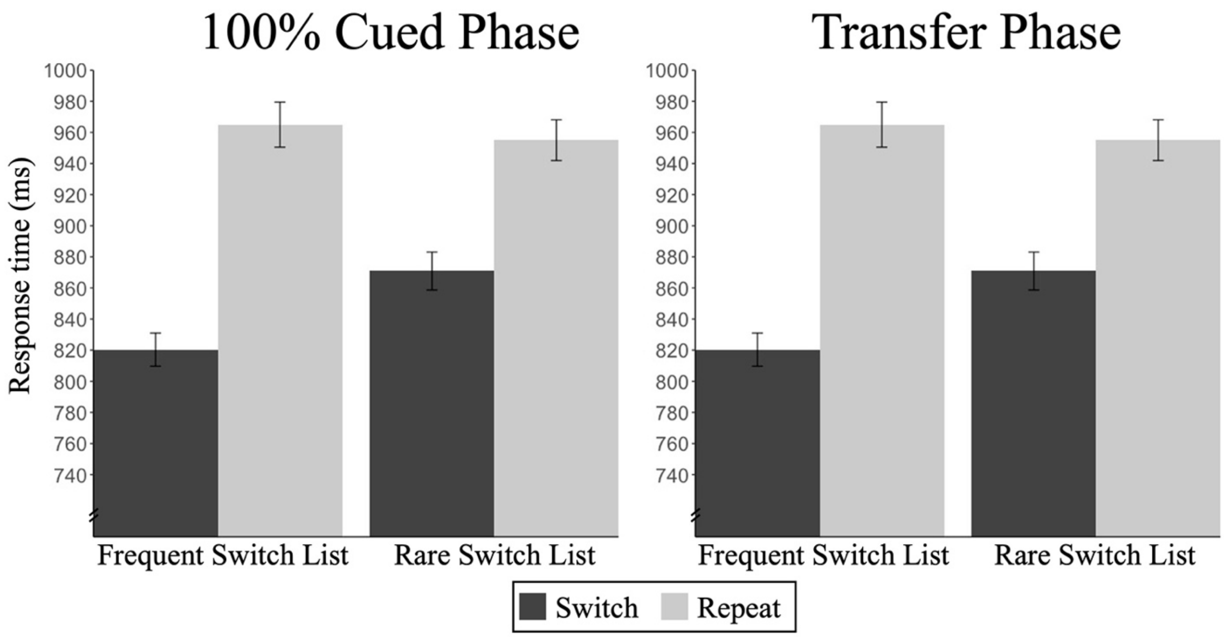
Only 1.16% of trials from each participant on average were removed as RT outliers. An additional 3.89% of trials on average were excluded from the RT and VSR analysis as errors or trials immediately following errors. After exclusions were applied, the mean error rate across all conditions was 8.56% (*SD* = 4.15) and the mean response time was 904 ms (*SD* = 77). Response time was not correlated with accuracy, $r(93) = -.18$, $p = .086$, $BF_{01} = 1.05 \pm 0.00\%$, indicating the absence of an overall speed-accuracy tradeoff. Means for RT and ER in all conditions are shown in Table 1.

Table 1. Condition Means and Standard Deviations for Experiment 1

Phase	Item Type	List	Response Time (ms)		Error Rate (percentage	
			Switch	Repeat	Switch	Repeat
100 % Cued	Inducer	Frequent Switch	949(89)	867(91)	9.35(6.52)	7.63(7.78)
		Rare Switch	963(107)	811(73)	10.85(8.95)	5.40(3.01)
	Diagnostic	Frequent Switch	961(94)	875(79)	10.48(6.73)	7.68(5.05)
		Rare Switch	967(96)	829(75)	10.34(6.10)	5.86(3.88)
Hybrid	Inducer	Frequent Switch	933(94)	881(99)	10.51(6.43)	8.53(8.26)
		Rare Switch	957(107)	826(81)	12.53(8.69)	7.15(4.65)
	Diagnostic	Frequent Switch	937(97)	866(84)	11.62(7.97)	8.36(6.02)
		Rare Switch	943(106)	839(87)	11.76(7.08)	7.35(5.07)

100% Cued Phase

Response time. Participants took longer to respond on switch trials (*M* = 960, *SD* = 97) than on repeat trials (*M* = 846, *SD* = 84), indicating the presence of a significant switch cost (main effect of transition), $F(1, 94) = 503.35$, $p < .001$, $\eta_p^2 = .84$. Participants also responded slower in the frequent switch list (*M* = 893, *SD* = 115) than the rare switch list (*M* = 913, *SD* = 98), as the main effect of list was significant, $F(1, 94) = 29.99$, $p < .001$, $\eta_p^2 = .24$. The manipulation succeeded in evoking an LWSP effect such that switch costs were smaller in the frequent switch list (*M* = 84, *SD* = 68) as compared to the rare switch list (*M* = 145, *SD* = 69), $F(1, 94) = 119.35$, $p < .001$, $\eta_p^2 = .56$. Switch and repeat trial responses times in each list are plotted in Figure 3.

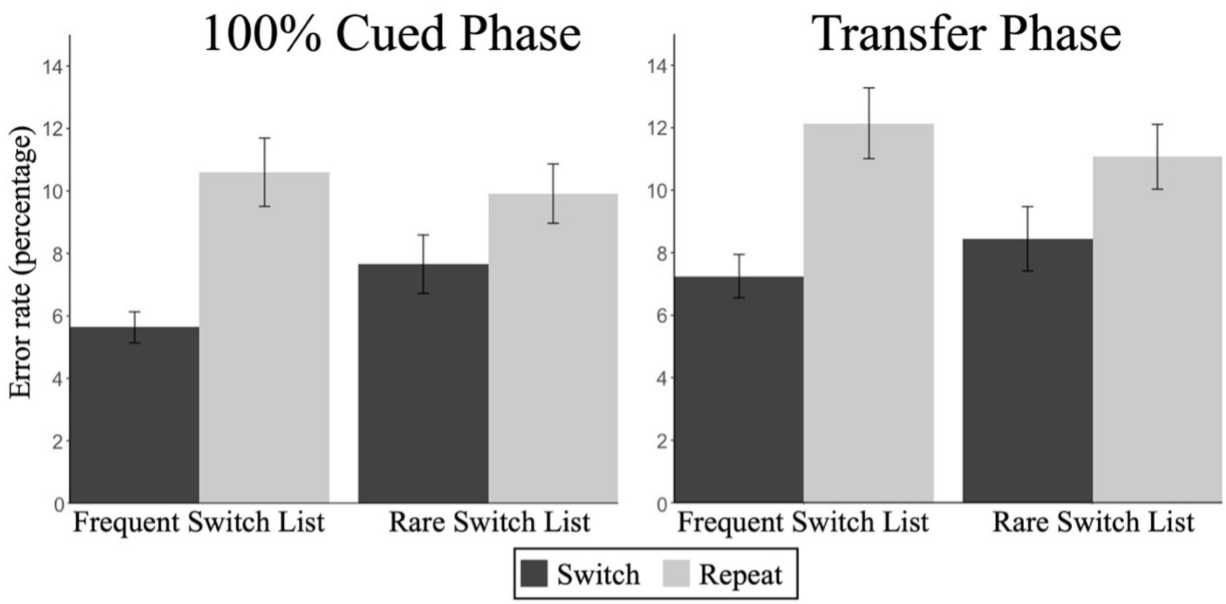


Note. Error bars represent within-subjects standard error of the mean (SEM).

Figure 3. Mean RT (in ms) as a function of list (frequent vs. rare switch) and transition (switch vs. repeat).

In addition to these effects, participants responded faster to inducer items (M = 897, SD = 110) than diagnostic items (M = 908, SD = 104), as indicated by the significant main effect of item type, $F(1, 94) = 12.76$, $p < .001$, $\eta_p^2 = .12$. The remaining interactions involving item type were not significant (item type x transition: $F(1, 94) = .90$, $p = .345$, $\eta_p^2 = .01$; list x item type: $F(1,94) = .02$, $p = .889$, $\eta_p^2 < .01$). Importantly, the item type x list x transition interaction was not significant, $F(1, 94) = 2.64$, $p = .107$, $\eta_p^2 = .03$, $BF_{01} = 4.91 \pm 2.9\%$, demonstrating that there was no difference in the LWSP effects between inducer and diagnostic items. The associated Bayes factor also indicated substantial evidence for a lack of an interaction effect, which serves as an important indicator of proactive metacontrol.

Error rate. Error rates were higher for switch trials (M = 10.26, SD = 7.15) than for repeat trials (M = 6.64, SD = 5.33); this switch cost was indicated by a significant main effect of transition, $F(1, 94) = 78.21$, $p < .001$, $\eta_p^2 = .45$ (Figure 4). The main effect of list was not significant, $F(1, 94) = 3.56$, $p = .062$, $\eta_p^2 = .04$. Similar to RT, there was a significant LWSP effect, as indicated by the significant list x transition interaction, $F(1, 94) = 15.33$, $p < .001$, $\eta_p^2 = .14$. Switch costs in ER were reduced in the frequent switch list (M = 80.29, SD = 62.36) relative to the rare switch list (M = 137.22, SD = 64.38).



Note. Error bars represent within-subjects SEM.

Figure 4. Mean ER (in %) as a function of list (frequent vs. rare switch) and transition (switch vs. repeat).

The main effect of item type was not significant, $F(1, 94) = .71$, $p = .403$, $\eta_p^2 = .01$. Nor did item type interact with the LWSP effect, $F(1, 94) = 2.15$, $p = .146$, $\eta_p^2 = .02$, $BF_{01} = 3.39 \pm 2.04\%$, again pointing to the presence of proactive metacontrol. Like in RT, item type also did not interact with the other variables (item type x transition: $F(1, 94) = .01$, $p = .938$, $\eta_p^2 < .01$; item type x list: $F(1, 94) = .92$, $p = .339$, $\eta_p^2 = .01$).

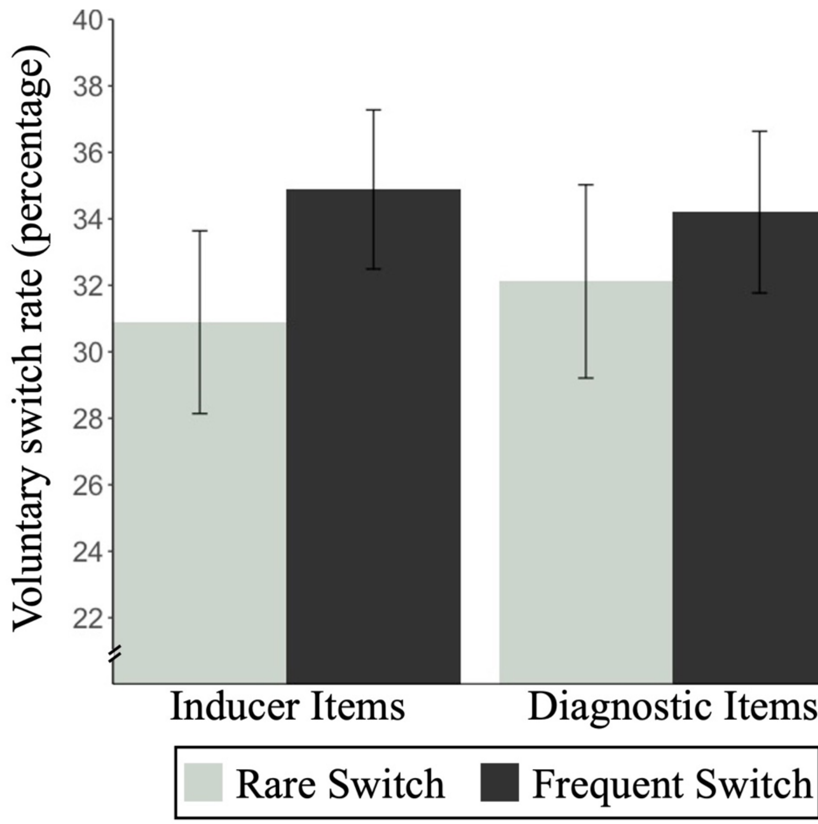
Hybrid Phase

Response time. As in the 100% cued phase, participants took longer to respond to switch trials (M = 943, SD = 101) than to repeat trials (M = 853, SD = 90), $F(1, 94) = 319.82$, $p < .001$, $\eta_p^2 = .77$. Likewise, participants took slightly longer to respond in the frequent switch list (M = 904, SD = 98) than in the rare switch list (M = 891.148, SD = 112.45), $F(1, 94) = 20.61$, $p < .001$, $\eta_p^2 = .18$. Importantly, switch costs were smaller when switches were frequent (M = 61, SD = 62) than when switches were rare (M = 118, SD = 69), indicating that the LWSP effect remained significant in the hybrid phase, $F(1, 94) = 77.30$, $p < .001$, $\eta_p^2 = .45$.

Unlike in the 100% cued phase, the main effect of item type was not significant, $F(1, 94) = 1.65$, $p = .202$, $\eta_p^2 = .02$. Also unlike the previous phase, item type interacted with the LWSP effect, $F(1, 94) = 26.07$, $p < .001$, $\eta_p^2 = .22$, $BF_{01} = 1.77 \pm 3.96\%$. A post-hoc ANOVA with only inducer items showed a significant LWSP effect of 81 ms (SD = 87), $F(1, 94) = 76.56$, $p < .001$, $\eta_p^2 = .45$. The other post-hoc analysis (with only diagnostic items) also showed a significant LSWP effect of 34 ms (SD = 62), $F(1, 94) = 27.48$, $p < .001$, $\eta_p^2 = .23$, supporting the presence of proactive metacontrol in the hybrid phase despite the significant 3-way interaction. The remaining interactions involving item type were not significant (transition x item type: $F(1, 94) = .94$, $p = .336$, $\eta_p^2 = .01$; list x item type: $F(1, 94) = 1.55$, $p = 216$, $\eta_p^2 = .02$.

Error rate. Switch costs remained significant, $F(1, 94) = 75.92$, $p < .001$, $\eta_p^2 = .45$, as participants made more errors on switch trials (M = 11.61, SD = 7.60) as compared to repeat trials (M = 7.85, SD = 6.16). The LSWP effect was significant like in the 100% cued phase, $F(1, 94) = 12.9$, $p < .001$, $\eta_p^2 = .12$, and did not differ between inducer and diagnostic items, $F(1, 94) = 2.78$, $p = .099$, $\eta_p^2 = .03$, $BF_{01} = 3.60 \pm 1.51\%$. The remaining effects were not significant (list: $F(1, 94) = .02$, $p = .877$, $\eta_p^2 < .01$; item type: $F(1, 94) = .09$, $p = .766$, $\eta_p^2 < .01$; transition x item type: $F(1, 94) = .06$, $p = .804$, $\eta_p^2 < .01$; list x item type: $F(1, 94) = 1.52$, $p = .221$, $\eta_p^2 = .02$).

Voluntary switch rate. VSR was higher in the frequent switch list (M = 34.54, SD = 11.82) than in the rare switch list, (M = 31.50, SD = 13.87), $F(1, 94) = 7.91$, $p = .006$, $\eta_p^2 = .08$ (Figure 5). The main effect of item type was not significant, $F(1, 94) = .07$, $p = .791$, $\eta_p^2 < .01$, indicating a similar overall VSR for inducer and diagnostic items. Importantly, there was no interaction between list and item type, $F(1, 94) = 1.01$, $p = .318$, $\eta_p^2 = .01$, $BF_{01} = 5.03 \pm .98\%$, indicating that the difference in VSR between the frequent and rare switch lists was not different between inducer and diagnostic items. Note that the BF indicates substantial evidence against this interaction, as the data are approximately 5 times more likely under a model without the interaction, which again support the presence of proactive metacontrol over free choice trials during the hybrid phase.



Note. Error bars represent within-subjects SEM.

Figure 5. VSR (in %) as a function of item type (inducer, diagnostic) and list (frequent switch, rare switch.

Discussion

The objective of this experiment was to determine whether a list-wide switch probability manipulation would concurrently modulate switch costs and VSR. The hypothesis that proactive metacontrol over switch readiness can modulate switch preferences as well as switch performance was supported. First, I found robust LWSP effects in both RT and in ER, replicating previous studies (e.g., Kang & Chiu, 2021). Second, I found a significant difference in VSR between lists, extending previous findings where VSR was modulated by item-specific switch probability associations (e.g., Chiu et al., 2020). Critical to my hypothesis, I did not find a difference between frequency-biased inducer items and frequency-unbiased diagnostic items either in the LWSP effects or in the VSR effect. Therefore, Experiment 1 addressed a research gap by demonstrating a list-wide, proactive modulation of VSR. Moreover, the inclusion of the VSR variable helped to validate the use of switch costs as a measure of learned switch readiness, as both measures followed a similar pattern. Follow-up studies of switch readiness, including Experiment 2, should be able to make valid claims about switch readiness on the basis of switch costs. In Experiment 2, I extended the findings of Experiment 1 and further examined the task-specificity of proactive metacontrol over switch readiness.

EXPERIMENT 2

Method

Participants

165 total participants were recruited (101 female, 57 male; mean age = 18.59, *SD* = 1.00), but 15 were excluded (4 due to poor performance, 10 due to having at least one condition cell with no valid trials after excluding outlier trials, and 1 for being below 18 years old). The following analyses and report were based on a sample of 150, which was a sample size estimated to provide 95% power for a small effect size based on means and standard deviations reported in an experiment with a somewhat similar design (Sabah et al., 2019–specifically the FC/VS. group and comparing between baseline and the second transfer block). The target sample size and the following experimental protocol were preregistered at aspredicted.org (https://aspredicted.org/7D5_8H2). Participants completed this Purdue IRB-approved study in exchange for credits towards a class requirement.

Apparatus and Stimuli

The experiment was presented in-person on a 21.5-inch Dell Flat Panel Monitor (different from Experiment 1) using PsychoPy version 2020.2.10 (Peirce et al., 2019). Stimuli consisted of 32 images of human-made objects (Possidónio et al., 2019). All images were piloted and confirmed to be easy to categorize as larger vs. smaller than a shoebox, primarily metal vs. nonmetal, and typically used/found indoors vs. outdoors. In addition, half of the images were tinted yellow, and the other half were tinted purple. Piloting confirmed that categorizations based on tint were of similar difficulty as categorizations along the other dimensions. A separate set of 16 images was used for practice exclusively. All dimensions were fully orthogonal. Images were 12.80 cm wide x 9.40 cm tall and were surrounded by a .30 cm thick rectangular frame that could be green, orange, red, or blue.

Design and Procedure

Participants performed a cued task switching procedure with each task being a different orthogonal dimension upon which an image could be dichotomously categorized. At any given time, participants would alternate between two out of four possible tasks: size – larger or smaller than a shoebox; material – primarily metal or nonmetal; location – primarily found/used indoors or outdoors; and color – tinted yellow or purple. Tasks were consistently paired across all participants; a phase of the experiment would always feature either the size and material tasks together or the location and color tasks. Participants were cued which task to perform on each trial by a colored rectangular frame that appeared around the image. For each participant, blue, orange, green, and red were mapped to different tasks.

Crucially for the research question regarding the generalizability of learned flexibility, the experiment was divided into two phases—a learning phase and a transfer phase. During the learning phase, participants alternated between lists with frequent and rare task switches to induce proactive metacontrol, just like in Experiment 1. Stimuli in one list never appeared in the other list, allowing participants to associate stimuli with a given list and a given switch probability. During the transfer phase, stimuli were presented according to the same list structure (one block of stimuli from each list) but switch probabilities for both lists were all 50%. One task pair was used in the learning phase, and the other was used in the transfer phase. Thus, the procedure induced a list-wide switch probability effect during the learning phase, and then assessed whether it persisted after replacing the tasks while presenting the same stimuli within each list. See Figure 6 for an overview of all experiment components.

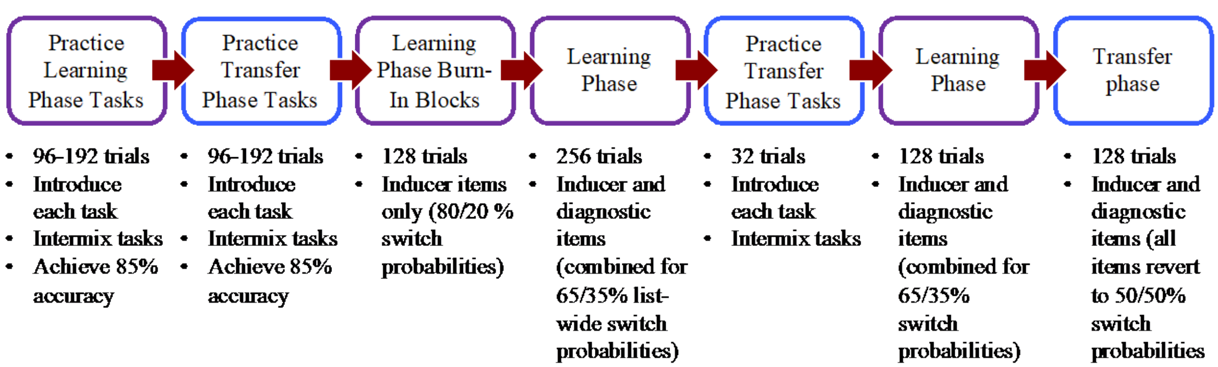


Figure 6. Experiment 2 design.

Participants performed 6 blocks of 64 trials in the learning phase, with inducer and diagnostic items intermixed at equal proportions. Across blocks, participants alternated between frequent and rare switch probability lists (i.e., rare-frequent-rare-frequent…), and the order was counterbalanced across participants. After the 4th block, participants briefly practiced the transfer tasks once more with 32 trials using the practice stimuli and a 50% switch probability. Next, participants completed the final two learning phase blocks just like the first four. The transfer phase immediately followed, and the list alternation pattern was preserved—if the last learning phase block was from the frequent switch list, the first transfer phase block would consist of items from the (formerly) rare switch block. As previously mentioned, participants now performed the other task pair and the switch probabilities in both lists became 50%. See Figure 6 for a breakdown of all experiment components.

Trials in both phases followed the same trial progression as Experiment 1. Before each block, participants were informed of their progress, "Block ___ of 11" (the brief practice block was included in this count, as were the burn-in blocks). They were also reminded of the correct responses for the upcoming task pair and allowed to pause for a break. If accuracy on the preceding block fell below 80%, participants were also asked to respond more accurately in the next block. After completing the transfer phase, participants provided demographic information and were granted credit towards their psychology class.

After providing informed consent, participants practiced all four tasks. Practice consisted of between 192 and 384 trials, with participants repeating blocks of 16 or 32 trials any time block accuracy fell below 85%. Participants first practiced the tasks that would occur in the learning phase, then the tasks that would occur in the transfer phase. A different set of stimuli was used for practice exclusively. Once participants became proficient in all four tasks, they began the learning phase of the main experiment. As in Experiment 1, the learning phase began with two 'burn-in' blocks (one from each list, presented in counterbalanced order across participants) with inducer but no diagnostic items. These burn-in blocks were not included in analysis.

Analysis

As in Experiment 1, analysis began by confirming manipulation success in terms of the switch probability effect by subjecting data (response time and error rate) from the learning phase to a 2 (item type: inducer, diagnostic) x 2 (list: frequent switch, rare switch) x 2 (transition: switch, repeat) within-subjects repeated measures ANOVA. To address the main question of flexibility transfer to entirely new task pairs, these same ANOVAs were performed on response times and error rates in the transfer phase.

In this experiment, ruling out WMU thresholds in favor of CA as the mechanism underlying LWSP effects required that there be no LWSP effect (list x transition interaction) in the transfer phase. To demonstrate evidence against transfer phase LWSP effects, I conducted JZS Bayes factor ANOVAs and report the BF_{01} for the critical list x transition interaction in the transfer phase. Using default priors on version 0.9.12-4.2 of the BayesFactor R package (Rouder et al., 2012), I computed a BF_{01} (H0/H1) to quantify evidence that the data came from a model (H0) with all effects in the design except for the LWSP effect vs. from a model (H1) with all effects included.

Results

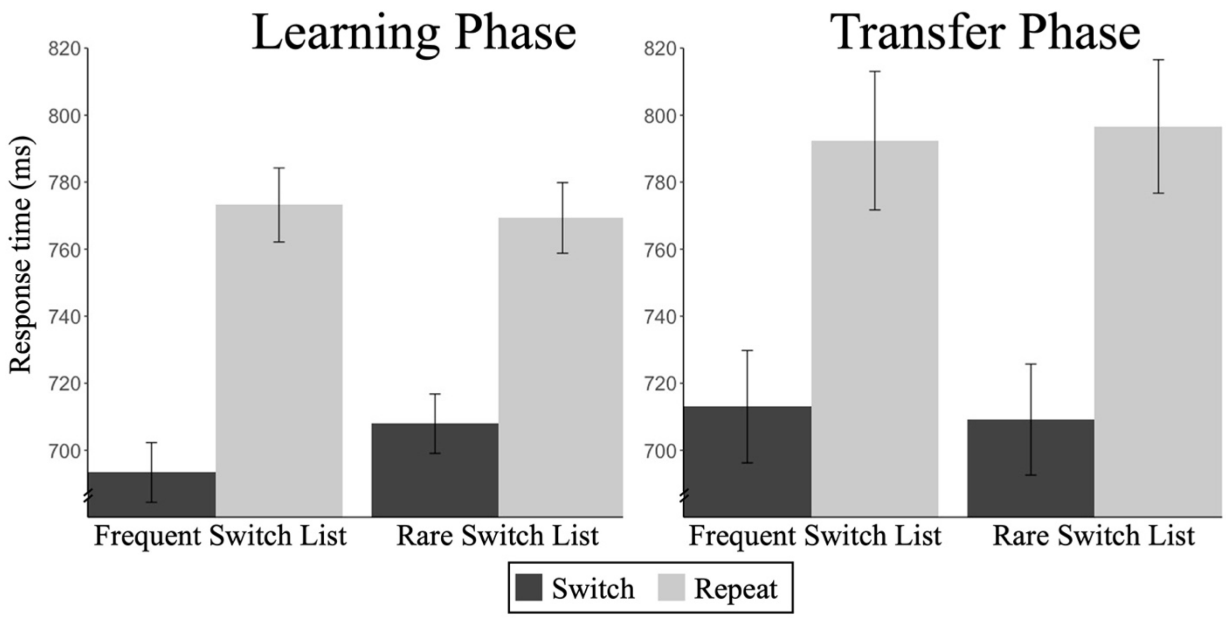
Very few trials from each participant (1.68% on average) were removed as RT outliers or as the first trials in a block. The mean error rate across all conditions was 28.25% (*SD* = 7.44%) and the mean response time was 739 ms (*SD* = 91). The increased error rate relative to Experiment 1 is perhaps not surprising given the difficulty of learning and applying four sets of task rules, even though only two were used at any given time. An additional 46.47% of trials on average were excluded from the RT analysis as errors or trials immediately following errors. Condition means for RT were estimated based on an average of 27.49 trials (*SD* = 12.68). While more than the ideal number of trials were excluded from RT analysis, the critical pattern of results (a significant LWSP effect in learning phase but not in the transfer phase) remained unchanged when trials following errors were included in the RT analysis rather than being excluded. This consistency supports the reliability of the mean estimates in the analyses reported here. Moreover, despite the complexity of the design, participants performed the appropriate task well above chance level as determined by samples drawn from a binomial distribution, indicating general understanding of the rules and response mappings. Since fewer trials were excluded from the ER analyses, mean estimates for ER were highly reliable and also agreed with the pattern of RT results. Therefore, the higher error rate should not preclude interpretations of RT LWSP effects. However, it did drop accuracy below ceiling, allowing for the detection of ER interactions that are often not found in similar studies (see section on response compatibility below). Response time was positively correlated with accuracy, $r(148) = .43$, $p < .001$, $BF_{01} = 2.28\text{e-}04 \pm 0.00\%$, such that participants who performed the task quickly were also likely to perform it accurately. RT and ER for all conditions are shown in Table 2.

Table 2. Condition Means and Standard Deviations for Experiment 2

Phase	Item Type	List	Response Time (ms)		Error Rate (percentage)	
			Switch	Repeat	Switch	Repeat
Learning	Inducer	Frequent Switch	761(92)	704(84)	28.8(10.69)	19.64(11.66)
		Rare Switch	768(96)	686(71)	32.6(14.14)	16.78(8.64)
	Diagnostic	Frequent Switch	777(92)	713(71)	32.87(9.80)	20.75(10.26)
		Rare Switch	778(98)	702(84)	32.21(11.47)	19.70(9.60)
Transfer	Inducer	Frequent Switch	791(186)	713(139)	44(14.22)	27.77(14.31)
		Rare Switch	789(181)	713(145)	41.69(14.24)	26.42(14.34)
	Diagnostic	Frequent Switch	801(159)	710(147)	42.17(14.83)	28.08(14.05)
		Rare Switch	796(182)	715(149)	43.25(13.26)	28.83(15.51)

Learning Phase

Response time. As expected, participants took longer to respond on switch trials (M = 771, SD = 95) than on repeat trials (M = 701 SD = 78), $F(1, 149) = 439.92$, $p < .001$, $\eta_p^2 = .75$ (Figure 7). The main effect of list was not significant, $F(1, 149) = 1.78$, $p = .184$, $\eta_p^2 = .01$. Critically, there was a significant LWSP effect, $F(1, 149) = 23.13$, $p < .001$, $\eta_p^2 = .13$. Switch costs were smaller in frequent switch lists (M = 61, SD = 60) than in rare switch lists (M = 80, SD = 58). Establishing this robust LWSP effect in the learning phase was a critical first step for determining whether it would later transfer to new tasks.

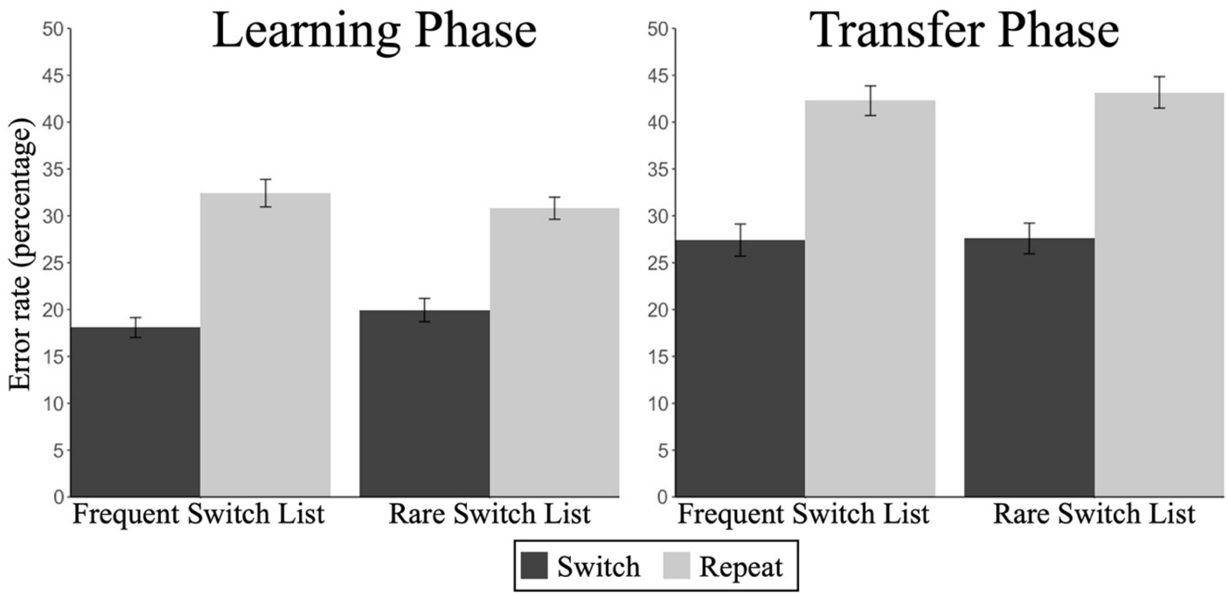


Note. Error bars represent within-subjects SEM.

Figure 7. Mean RT (in ms) as a function of list (frequent vs. rare switch) and transition (switch vs. repeat).

Participants also responded faster for inducer items (M = 729, SD = 93) than for diagnostic items (M = 742, SD = 94), as the main effect of item type was significant, $F(1, 149) = 40.84$, $p < .001$, $\eta_p^2 = .22$. Item type did not interact with transition, $F(1, 149) = .25$, $p = .617$, $\eta_p^2 < .01$, or list, $F(1, 149) = .04$, $p = .844$, $\eta_p^2 < .01$. Finally, the item type x list x transition 3-way interaction was not significant, $F(1, 149) = 1.80$, $p = .182$, $\eta_p^2 = .01$, suggesting that the LWSP effect was not different between inducer and diagnostic items.

Error rate. Along with the RT switch cost, participants made more errors on switch trials (M = 31.62, SD = 11.74) than on repeat trials (M = 19.01, SD = 10.19), $F(1, 149) = 571.87$, $p < .001$, $\eta_p^2 = .79$ (Figure 8). The variable list was not significant, $F(1, 149) = .10$, $p = .754$, $\eta_p^2 = .01$, but the LWSP effect was significant, $F(1, 149) = 19.52$, $p < .001$, $\eta_p^2 = .12$. Switch costs were smaller when switches were frequent (M = 10.86, SD = 9.31) than when they were rare (M = 14.34, SD = 11.29).



Note. Error bars represent within-subjects SEM.

Figure 8. Mean ER (in %) as a function of list (frequent vs. rare switch) and transition (switch vs. repeat).

Following the same pattern as RT, participants made fewer errors on inducer items (*M* = 24.35, *SD* = 13.16) than on diagnostic items (*M* = 26.28, *SD* = 12.08), as indicated by a significant main effect of item type, $F(1, 149) = 22.09$, $p < .001$, $\eta_p^2 = .13$. This was further qualified by a significant item type x list x transition interaction whereby the LSWP effect differed between inducer and diagnostic items, $F(1, 149) = 18.57$, $p < .001$, $\eta_p^2 = .11$. Post-hoc analysis revealed a significant LWSP effect among inducer items (*M* = 6.62, *SD* = 14.30) $F(1, 149) = 33.87$, $p < .001$, $\eta_p^2 = .18$, but not among diagnostic items (*M* = .38, *SD* = 11.86), $F(1, 149) = .17$, $p = .683$, $\eta_p^2 < .01$. The remaining interactions were not significant (item type x transition: $F(1, 149) = .02$, $p = .878$, $\eta_p^2 < .01$; item type x list: $F(1, 149) = 2.96$, $p = .087$, $\eta_p^2 = .02$).

Transfer Phase

Response time. The switch cost remained significant in the transfer phase as response times were longer on switch trials (*M* = 794, *SD* = 178) than on repeat trials (*M* = 711, *SD* = 146), $F(1, 149) = 180.93$, $p < .001$, $\eta_p^2 = .55$ (Figure 7). The main effect of list was not significant, $F(1, 149) < .01$, $p = .983$, $\eta_p^2 < .01$. Critical to the question of whether CA contributes to learned switch readiness, the LWSP effect was not significant in the transfer phase, $F(1, 149) = 1.08$, $p = .301$, $\eta_p^2 = .01$, $BF_{01} = 10.77 \pm 3.69\%$. Note that the BF indicates the data are almost 11 times more likely to have come from distributions without an LWSP effect than one with such an effect, constituting "strong" evidence for the null according to Jeffreys (1961).

All remining effects were not significant (item type: $F(1, 149) = 1.01$, $p = .316$, $\eta_p^2 = .01$; item type x transition: $F(1, 149) = 1.50$, $p = .223$, $\eta_p^2 = .01$; item type x list: $F(1, 149) = .01$, $p = .904$, $\eta_p^2 < .01$; item type x list x transition: $F(1, 149) = .48$, $p = .492$, $\eta_p^2 < .01$).

Error rate. Participants continued to have higher error rates on switch trials (*M* = 42.72, *SD* = 14.36) than on repeat trials (*M* = 27.49, *SD* = 14.78) in the transfer phase, $F(1, 149) = 354.61$, $p < .001$, $\eta_p^2 = .70$ (Figure 8). As predicted, the list x transition interaction was not significant, $F(1, 149) = .30$, $p = .586$, $\eta_p^2 < .01$, $BF_{01} = 10.94 \pm 3.09\%$, again indicating no transfer. Similar to the LWSP effect finding in RT, the Bayes factor indicated strong evidence against a difference in switch costs between lists.

While the main effect of both list, $F(1,149) = .36$, $p = .547$, $\eta_p^2 < .01$, and item type $F(1,149) = .83$, $p = .365$, $\eta_p^2 = .01$, were not significant, the interaction between these variables was, $F(1,149) = 4.76$, $p = .031$, $\eta_p^2 = .03$. Post-hoc analysis revealed that the main effect of list was nonsignificant both among inducer items, $F(1,149) = 2.81$, $p = .096$, $\eta_p^2 = .02$, and among diagnostic items, $F(1,149) = .65$, $p = .42$, $\eta_p^2 < 01$. The remaining interactions were not significant (item type x transition: $F(1,149) = 1.46$, $p = .229$, $\eta_p^2 = .01$; item type x list x transition: $F(1,149) = .19$, $p = .662$, $\eta_p^2 < .01$).

Control Analyses

To rule out two confounds that might have contributed to the failure to detect the LWSP effects in the transfer phase, I conducted a series of post-hoc 'control' analyses. First, the transfer phase might have been too short to detect LWSP effects. Second, the LWSP effect might have become extinct gradually in the transfer phase due to all lists reverting to a 50% switch probability. To address the first concern, RT and ER from only the first two (non-burn-in) blocks of the learning phase were subjected to a 2 (item type: inducer, diagnostic) x 2 (list: frequent switch, rare switch) x 2 (transition: switch, repeat) within-subjects repeated measures ANOVA (the same analysis as the full learning phase and transfer phase) to determine whether two blocks were sufficient to detect LWSP effects.

To address the second concern, two analysis strategies were used. First, both transfer phase blocks were divided into first and second halves, and the variable half was added to the ANOVA as an additional within-subjects factor. This allowed for investigation as to whether the LWSP effect was stronger in the first half of each list compared to the second half, which would support the presence of a gradual extinction. A drawback of this analysis is that several participants had to be excluded from the RT analysis for having no valid trials in at least one of the conditions, reducing the sample size to 131. Second, I conducted a pair of mixed model ANOVAs for RT and ER with list (frequent switch, rare switch) as a between-subjects variable but transition and item type as within-subjects variables. In this analysis, I only included the first block of the transfer phase, which was counterbalanced across participants to present either the (formerly) frequent switch or rare switch list. This allowed for a finer grain look at the LWSP effects in the early portion of the transfer phase. For both analyses addressing the concern of extinction, I also present the BF_{01} for the critical interactions of interest.

Transfer Phase Too Short?

Response time. Since the transfer phase consisted of only two blocks, a control analysis with a comparable amount of data from the learning phase still showed a significant LWSP effect (M = 16, SD = 113), $F(1, 149) = 6.41$, $p = .012$, $\eta_p^2 = .04$, indicating the shorter transfer phase did not preclude detection of the potential transfer effect. One difference between this smaller dataset and the learning phase analysis was that the 3-way interaction between item type, list, and transition was significant, $F(1, 149) = 6.64$, $p = .011$, $\eta_p^2 = .04$. This was driven by a larger LWSP effect among inducer items (M = 31, SD = 112) than among diagnostic items (M = -1, SD = 108).

All other effects of this limited-dataset analysis mirrored the pattern of the main learning phase ANOVA (transition: $F(1, 149) = 280.35$, $p < .001$, $\eta_p^2 = .65$; list: $F(1, 149) = .63$, $p = .429$, $\eta_p^2 < .01$; item type: $F(1, 149) = 35.66$, $p < .001$, $\eta_p^2 = .19$; item type x transition: $F(1, 149) = .21$, $p = .651$, $\eta_p^2 < .01$; item type x list: $F(1, 149) = .10$, $p = .758$, $\eta_p^2 < .01$)

Error rate. A learning phase LWSP effect of 3.07% (*SD* = 23.29) remained detectable when only two blocks were included, $F(1, 149) = 4.53$, $p = .035$, $\eta_p^2 = .03$, indicating that the length of the transfer phase did not preclude detection. The pattern of results for all other effects matched the original ANOVA (transition: $F(1, 149) = 316.28$, $p < .001$, $\eta_p^2 = .68$; list: $F(1, 149) = 1.08$, $p = .301$, $\eta_p^2 = .01$; item type: $F(1, 149) = 28.94$, $p < .001$, $\eta_p^2 = .16$; item type x transition: $F(1, 149) = .16$, $p = .685$, $\eta_p^2 < .01$; item type x list: $F(1, 149) = 1.08$, $p = .300$, $\eta_p^2 = .01$; item type x list x transition: $F(1, 149) = 13.04$, $p < .001$, $\eta_p^2 = .08$).

Possibility of Extinction?

Response time. Two follow up analyses indicated that the LWSP effect was undetectable immediately in the transfer phase, rather than gradually diminishing as a result of the 50% switch probability. First, when the transfer phase blocks were divided into halves, participants responded faster in the second half of both lists (*M* = 752, *SD* = 170) than in the first half (*M* = 759, *SD* = 181), $F(1, 130) = 10.14$, $p = .002$, $\eta_p^2 = .07$. This experience effect was similar between frequent and rare switch lists, $F(1, 130) = 3.85$, $p = .052$, $\eta_p^2 = .03$. Switch costs remained significant, $F(1, 130) = 196.17$, $p < .001$, $\eta_p^2 = .60$, while the list x transition interaction remained nonsignificant, $F(1, 130) = .57$, $p = .451$, $\eta_p^2 < .01$, $BF_{01} = 13.39 \pm 5.3\%$. Moreover, there was no difference in the size of the list x transition interaction effect in the first vs. second half of each list, $F(1, 130) = .14$, $p = .704$, $\eta_p^2 < .01$, $BF_{01} = 10.06 \pm 6.33\%$. The Bayes factor indicated strong evidence for the null model, suggesting that the LWSP effect disappeared quickly once the transfer phase started.

All other effects, including any interactions with half, were not significant (list: $F(1, 130) = .20$, $p = .654$, $\eta_p^2 < .01$; item type: $F(1, 130) = .94$, $p = .334$, $\eta_p^2 = .01$; item type x transition: $F(1, 130) = .24$, $p = .625$, $\eta_p^2 < .01$; item type x list: $F(1, 130) = .18$, $p = .675$, $\eta_p^2 < .01$; transition x half: $F(1, 130) = .24$, $p = .624$, $\eta_p^2 < .01$; item type x half: $F(1, 130) = .11$, $p = .74$, $\eta_p^2 < .01$; item type x list x transition: $F(1, 130) = .69$, $p = .409$, $\eta_p^2 < .01$; item type x transition x half: *F*(1,

130) = .21, p = .647, η_p^2 < .01; item type x list x half: $F(1, 130)$ = .10, p = .752, η_p^2 < .01; item type x list x transition x half: $F(1, 130)$ = 1.25, p = .266, η_p^2 = .01).

Similarly, when analyzing only the first block of the transfer phase as a between-subjects variable, the LWSP effect remained nonsignificant, $F(1, 148)$ = 1.04, p = .310, η_p^2 = .01, BF_{01} = 6.65 ± 3.43%. The Bayes factor indicated substantial evidence against a difference in switch costs between participants completing different lists. The main effect of transition remained significant as in the original analysis, $F(1, 148)$ = 114.28, p < .001, η_p^2 = .44, while all other effects remained nonsignificant (list: $F(1, 148)$ = .17, p = .682, η_p^2 < .01; item type: $F(1, 148)$ = .23, p = .632, η_p^2 < .01; item type x transition: $F(1, 148)$ = 1.97, p = .163, η_p^2 = .01; item type x list: $F(1, 148)$ = .62, p = .433, η_p^2 < .01; item type x list x transition: $F(1, 148)$ = .94, p = .333, η_p^2 = .01).

Error rate. Similar to RT, I failed to detect the LWSP effect in the control analyses with ER. When half was added to the transfer phase ANOVA as a within-subjects factor, participants performed significantly better in the second half (*M* = 34.78, *SD* = 19.72) than the first half (*M* = 37.50, *SD* = 20.25), $F(1, 149)$ = 16.07, p < .001, η_p^2 = .10. The list x transition effect remained non-significant, $F(1, 149)$ = .08, p = .776, η_p^2 < .01, BF_{01} = 13.84 ± 5.21%. The Bayes factor again indicated strong evidence against this interaction. Critically, the list x transition interaction effect did not interact with half, $F(1, 149)$ = 1.30, p = .256, η_p^2 = .01, BF_{01} = 8.15 ± 8.42%, and the Bayes factor revealed substantial evidence that the ER LWSP effect was negligible in both halves of the transfer blocks rather than diminishing from the first half to the second.

Only one effect, item type, interacted with half, $F(1, 149)$ = 5.75, p = .018, η_p^2 = .04, which was due to the difference between halves being larger among diagnostic items (*M* = 4.26, *SD* = 21.36) than among inducer items (*M* = 1.485, *SD* = 22.08). No other effects interacted with half (transition x half: $F(1, 149)$ = .34, p = .558, η_p^2 < .01; list x half: $F(1, 149)$ = .62, p = .434, η_p^2 < .01; item type x transition x half: $F(1, 149)$ = 1.34, p = .248, η_p^2 = .01; item type x list x half: $F(1, 149)$ = .01, p = .902, η_p^2 < .01; item type x list x transition x half: $F(1, 149)$ = 1.22, p = .271, η_p^2 = .01). Unlike the original transfer phase ANOVA, list interacted with item type, $F(1, 149)$ = 4.47, p = .036, η_p^2 = .03, such that the difference between lists was larger among diagnostic items (*M* = 4.26, *SD* = 21.40) than among inducer items (*M* = 1.49, *SD* = 22.08).

The switch cost remained significant, $F(1, 149)$ = 334.09, p < .001, η_p^2 = .69, and all other effects remained non-significant (list: $F(1, 149)$ = .01, p = .931, η_p^2 < .01; item type: $F(1, 149)$ =

1.19, p = .278, η_p^2 = .01; item type x transition: $F(1, 149)$ = .86, p = .356, η_p^2 = .01; item type x list x transition: $F(1, 149)$ = .82, p = .366, η_p^2 < .01).

Likewise, a second analysis that only considered the first block in the transfer phase and treated the switch probability of this block as a between-subjects factor was unable to detect a list x transition interaction effect, $F(1, 148)$ = .78, p = .379, η_p^2 = .01, BF_{01} = 6.09 ± 3.34. The Bayes factor suggests substantial evidence against the effect.

Switch costs remained significant, $F(1, 148)$ = 231.52, p < .001, η_p^2 = .61, as did the item type x list interaction, $F(1, 148)$ = 4.00, p = .047, η_p^2 = .03. All other effects were not significant (list: $F(1, 148)$ < .01, p = .971, η_p^2 < .01; item type: $F(1, 148)$ = .61, p = .435, η_p^2 < .01; item type x transition: $F(1, 148)$ = .36, p = .552, η_p^2 < .01; item type x list x transition: $F(1, 148)$ = .01, p = .931, η_p^2 < .01).

Response Compatibility and Mediated Pathways

The design also allowed me to measure differences in RT and ER based on response compatibility (whether correct responses for both tasks for a given stimulus are mapped to the same or different keys). The size of the response compatibility effect has been thought of as a marker of task-set shielding and another way to index cognitive flexibility (e.g., Fischer & Hommel, 2012; Janczyk, 2016; Plessow et al., 2012; Zwosta et al., 2013). A possible consequence of increased CA is that stimuli are increasingly categorized according to both tasks even though only one is required. Such inefficient dual tasking has been proposed as a mechanism behind response compatibility effects because evidence thresholds in favor of a particular response (e.g., ‘press the left key’) should be met quicker when both categorization rules accumulate evidence for the same response than when they do not. This is referred to as a mediated route to response, as the processing sequence from stimulus onset to final response requires a mediating step of representing the stimulus according to all task categories (e.g., Kiesel et al., 2007; Schneider, 2015, 2018). The mediated route continues to produce response compatibility effects when alternative sources of the effect have been ruled out (Schneider, 2015), and such dual-task processing seems difficult to avoid despite the performance impediment on incompatible trials. I investigated whether learned switch readiness would increase response compatibility effects, a potential consequence of the CA mechanism. This was done using a 2 (compatibility: compatible, incompatible) x 2 (item type: inducer, diagnostic) x 2 (list: frequent switch, rare switch) repeated measures ANOVA. The primary effect of interest was the interaction between compatibility and list, especially whether compatibility effects become larger in the frequent switch list than in the rare switch list. Such an interaction would both serve as an additional index of cognitive flexibility and also point to a consequence of CA that could conceivably not be present for WMU-based flexibility. For simplicity, compatibility was only examined relative to the two tasks being completed in a given phase, even though the 4-task structure conceivably allowed for several combinations of compatibility or compatibility interference across phases.

Learning Phase

Response time. As expected, incompatible trials with opposing correct responses for the two tasks incurred longer response times (M = 760, SD = 86) than compatible trials with the same correct response for either task (M = 710, SD = 86), $F(1, 149) = 278.38$, $p < .001$, $\eta_p^2 = .65$. However, the compatibility effect did not differ between frequent and rare switch lists, suggesting that learned switch readiness might not have modulated the extent to which stimuli are categorized according to the irrelevant task rules, $F(1, 149) = 1.55$, $p = .215$, $\eta_p^2 = .01$. Although the response compatibility effect was significantly larger among inducer items (M = 57, SD = 55) than among diagnostic items (M = 42, SD = 55), $F(1, 149) = 14.14$, $p < .001$, $\eta_p^2 = .09$, the 3-way interaction between compatibility, item type, and list was not significant, $F(1, 149) = 2.03$, $p = 0.156$, $\eta_p^2 = .01$.

Response times generally differed between the lists, with participants taking longer to respond in the frequent switch list (M = 723, SD = 90) than in the rare switch list (M = 747, SD = 88), $F(1, 149) = 38.8$, $p < .001$, $\eta_p^2 = .21$. Response times also differed by item type, with participants taking longer to respond to diagnostic items (M = 742, SD = 88) than to inducer items (M = 728, SD = 90), $F(1, 149) = 66.07$, $p < .001$, $\eta_p^2 = 0.31$. These two variables interacted, $F(1, 149) = 71.08$, $p < .001$, $\eta_p^2 = .32$, which was due to a larger list effect (frequent vs. rare switch) among inducer items (M = 43.45, SD = 67.68) than among diagnostic items (M = 5.62, SD = 63.28).

Error rate. As mentioned earlier, the higher ER enabled me to observe several significant main effects and interactions in this analysis. Participants made significantly more errors on incompatible trials (*M* = 33.91, *SD* = 11.41) than on compatible ones (*M* = 16.09, *SD* = 11.12), $F(1, 149) = 679.37$, $p < .001$, $\eta_p^2 = .82$. They also made more errors in the frequent switch list (*M* = 26.66, *SD* = 14.88) than in the rare switch list (*M* = 23.34, *SD* = 13.64), $F(1, 149) = 61.59$, $p < .001$, $\eta_p^2 = .29$. Likewise, they made more errors when responding to diagnostic items (*M* = 26.39, *SD* = 13.69) compared to inducer items (*M* = 23.62, *SD* = 14.89), $F(1, 149) = 55.73$, $p < .001$, $\eta_p^2 = .27$. These main effects were further qualified by interactions. The compatibility effect was larger among inducer items (*M* = 18.63, *SD* = 12.03) than among diagnostic items (*M* = 17.15, *SD* = 11.05), $F(1, 149) = 4.36$, $p = .038$, $\eta_p^2 = .03$. List also interacted with item type, $F(1, 149) = 38.25$, $p < .001$, $\eta_p^2 = .20$, such that the difference between lists was larger among inducer items (*M* = 5.77, *SD* = 10.04) than among diagnostic items (*M* = .89, *SD* = 9.47).

The response compatibility effect was larger in the frequent switch list (*M* = 19.02, *SD* = 12.58) than in the rare switch list (*M* = 16.74, *SD* = 10.39), $F(1, 149) = 9.27$, $p = .003$, $\eta_p^2 = .06$. This finding suggests increased relaxing of task-set shielding, perhaps as a consequence of CA-based flexibility. The difference in size of the response compatibility effect between the two lists was larger for inducer items (*M* = 4.42, *SD* = 12.93) than for diagnostic items (*M* = .21, *SD* = 13.19), as evidenced by a significant 3-way interaction, $F(1, 149) = 7.94$, $p = .005$, $\eta_p^2 = .05$.

Transfer Phase

Response time. This particular analysis was run with fewer participants due to missing data in some conditions. Like the learning phase, participants took longer to respond to incompatible trials (*M* = 753, *SD* = 158) than to compatible trials (*M* = 744, *SD* = 155), $F(1, 146) = 6.57$, $p = .011$, $\eta_p^2 = .04$. Once more, this response compatibility effect did not differ between lists, $F(1, 146) < .01$, $p = .968$, $\eta_p^2 < .01$. All other effect were also nonsignificant (list: $F(1, 146) < .01$, $p = .970$, $\eta_p^2 < .01$; item type: $F(1, 146) = 2.16$, $p = .143$, $\eta_p^2 = .01$; compatibility x item type: $F(1, 146) = .21$, $p = .650$, $\eta_p^2 < .01$; list x item type: $F(1, 146 = .20$, $p = .655$, $\eta_p^2 <.01$; compatibility x item type x list: $F(1, 146) = .25$, $p = .616$; $\eta_p^2 < .01$).

Error rate. As before, participants made more errors on incompatible trials (M = 40.49, SD = 14.32) than compatible ones (M = 30.07, SD = 14.72), $F(1, 149) = 174.63$, $p < .001$, η_p^2 = .54, and this compatibility effect did not differ between the two lists, $F(1, 149) = .05$, $p = .827$, $\eta_p^2 < .01$. While both the main effects of list ($F(1, 149) = .27$, $p = .602$, $\eta_p^2 < .01$) and item type ($F(1, 149) = .77$, $p = .381$, $\eta_p^2 < .01$) were not significant, these variables interacted $F(1, 149) = 3.99$, $p = .048$, $\eta_p^2 = .03$. However, post-hoc analysis revealed no significant effect of list either among the inducer items (M = 2, SD = 18), $F(1, 149) = 2.81$, $p = .096$, $\eta_p^2 = .02$, or among the diagnostic items (M = -1, SD = 17), $F(1, 149) = .65$, $p = .420$, $\eta_p^2 < .01$. All remaining effects were not significant (compatibility x item type: $F(1, 149) = .01$, $p = .999$, $\eta_p^2 < .01$; compatibility x item type x list: $F(1, 149) = .59$, $p = .445$, $\eta_p^2 < .01$).

Discussion

The goal of this experiment was to investigate whether lists of stimuli associated with frequent task switches would maintain increased switch readiness after the tasks learned in a task-switching paradigm were replaced with novel tasks. Specifically, based on the hypothesis that learned switch readiness is primarily achieved through a CA mechanism, I expected that replacing the tasks used to build an LWSP effect with two new tasks while leaving stimuli unchanged would eliminate the LWSP effect. In support of my hypothesis, the results showed a robust LWSP effect in the learning phase that was eliminated in the transfer phase. Additional control analyses further ruled out possible confounds such as not enough trials in the transfer phase or gradual extinction caused by unbiased switch probabilities. The lack of LWSP effects in the transfer phase indicates that the learned switch readiness elicited by the switch probability manipulation is highly task specific. Such task-specificity fits well with the concurrent activation (CA) mechanism of switch readiness proposed by Dreisbach and Fröber (2019).

GENERAL DISCUSSION

In two highly powered experiments, I sought to determine whether list-wide switch probability effects created in one context would generalize to a different context. In Experiment 1, the learning context was a 100% cued phase where flexible behavior was entirely involuntary, and the dissimilar context was a hybrid phase including trials where flexible behavior was voluntary. In Experiment 2, the learning context consisted of one pair of tasks, and the dissimilar context consisted of a transfer phase with new tasks without a list-wide switch probability manipulation.

Experiment 1 demonstrated that LWSP effects could be detected by measuring VSR, extending previous findings that item-specific manipulations could also modulate VSR. Thus, it addressed an unanswered question as to whether proactive metacontrol, like reactive metacontrol, can be triggered bottom-up to influence the decision to repeat or switch a task. Moreover, these results underscored the importance of associative learning and automatic retrieval of control states in cognitive control.

While contributing to our understanding of learned switch readiness in its own right, Experiment 1 also served as an assumption check for Experiment 2. Experiment 1 demonstrated that learned flexibility can generalize beyond the context in which it was learned, so long as the generalization is within-task (i.e., the same tasks are used in both contexts). This suggests that Experiment 2, with its similar learning phase, should have been sufficient to detect transfer effects should they be present. Moreover, the agreement between switch cost and VSR LWSP effects in Experiment 1 lend credence to the claim that decreasing switch costs index increased switch readiness (see also Ravizza & Carter, 2008), and that such switch readiness is what was manipulated in the current paradigm. Accordingly, the results of Experiment 2 indicated that the learned associations between a list context and a high demand for switching that constitute learned switch readiness are task-specific, and will not support efficient switching among new tasks, despite the presence of cues previously associated with differential switch demands.

Such task-specificity is consistent with the concurrent activation (CA) mechanism proposed to support demand-based modulations in switch costs (Dreisbach & Fröber, 2019). According to this proposal, contexts in which switches are frequent will cause both sets of task rules to remain active in working memory, even though only one is needed at a given time. To speculate, this change could be reflected in existing computational models either by a) decreasing competition parameters between the two tasks such that more activation for one task does not necessarily mean less activation for the other, or b) decreasing task-set inhibition (Koch et al., 2010; Schneider, 2007) such that a task ceases to be actively inhibited when it is switched away form. The former possibility is a feature of the Metacontrol State Model (Hommel, 2015). This model conceptualizes flexible behavior as two choices which compete for activation, and one of the choices is more goal-appropriate and thereby receives extra activation via a top-down bias. In this model, either relaxation of the top-down bias or decreasing competition between the two alternatives results in more flexible behavior. On the other hand, the latter possibility of relaxed task-set inhibition could be evaluated via a switching paradigm using 3 tasks (Mayr & Keele, 2000). In such a paradigm, a typical finding is that it is more difficult to switch back to a task that had been switched away from in the previous trials than it is to switch to a third task. This is referred to as n-2 repetition costs. To put it concretely, when presented with 3 possible tasks (Task A, Task B, and Task C), the third trial in a sequence of 3 consecutive trials will incur a performance cost if the trial sequence is ABA rather than ABC (Arbuthnott & Frank, 2000; Lien & Ruthruff, 2008; Mayr & Keele, 2000; Philipp et al., 2007) This effect has been interpreted as indicating that a task is normally inhibited when it is switched away from (e.g., Mayr & Kliegl, 2003, Schneider, 2007). If CA is driven by changes in task-set inhibition, the n-2 effect should diminish when switches are more frequent. While some researchers have investigated learned switch readiness among 3 tasks (Chiu & Egner; 2017, Siqi-Liu & Egner, 2020), additional insight into the impacts of switch frequency on task-set inhibition would be beneficial for further clarifying the mechanisms behind CA.

Regardless of how CA is instantiated, the results reported here compliment a recent study addressing a similar question. Siqi-Liu and Egner (2020) devised an LWSP paradigm like the learning phases of the paradigm presented here, but with three tasks to switch between instead of two. Critically, the LWSP manipulation was created among only two of the three tasks, with the third task equally likely to represent a task switch or repeat. LWSP effects were detected among trials with the biased tasks, but not unbiased task. These findings are easily interpreted through the CA framework: participants were keeping rules for the two biased tasks available during the frequent switch list but were not inclined to do so for the third set of task rules because it was not frequently needed for switches. Presumably, concurrent activation among three tasks is possible when demanded by a goal or context. Taken together, these studies begin to indicate convergent evidence for the task-specificity of learned switch readiness, with the present experiment taking the additional step of training flexibility in one context and testing it in a different one.

In Experiment 2, I also sought to determine whether learned switch readiness would lead to larger response compatibility effects, as increases in CA might result in relaxation of task-set shielding and increased interference from irrelevant task rules via the mediated pathway (Schneider, 2015). Accordingly, I expected a pattern of results whereby response compatibility effects differed between lists in the learning phase but not transfer phase. Contrary to expectations, the LWSP manipulation in the learning phase did not differ between lists for RT. The expected pattern was found, however, only for ER. One interpretation of this finding is that keeping both sets of task rules active increases the degree to which stimuli are simultaneously categorized according to both tasks, leading to larger compatibility effects through the mediated pathway. Thus, the compatibility effect in ER is also supportive of the CA mechanism. Because the response compatibility effects reported here represent the influence of both mediated and non-mediated routes, it is possible that the impact of non-mediated response compatibility obscured the effects of CA on mediated response compatibility for RT. Follow up studies wherein only one pathway is available (Schneider, 2015) could help clarify the mixed results.

An alternative to CA for explaining LWSP effects is that instead of keeping both sets of task rules active, participants respond to the frequent switch list by defaulting to task switches on all trials and switching a second time back to the appropriate task on repeat trials. While the current study was not designed to differentiate between these possibilities, the fact that repeat trial response times were numerically different between the two learning phase lists in Experiment 1 (47 ms difference) but not in Experiment 2 (less than 1 ms difference) hints that inappropriate switching is at least partially responsible for the LWSP effect in Experiment 1. This is because increases in concurrent activation should not necessarily cause unique harm to repeat trial performance, while inappropriate switching on repeat trials should. However, note that these differences were not subjected to significance testing. Future studies aimed at distinguishing learned flexibility in the form of CA vs. defaulting to switching would be beneficial.

A limitation particular to Experiment 2 is the concern that the 50% switch probability in the transfer phase—and not the new task pair—is what eliminated the LWSP. Multiple control analyses suggested that this was not the case, and that LWSP effects were undetectable very early into the transfer phase, before participants could fully learn that all items had new switch probabilities.

CONCLUSION

Learned associations between a temporal context and demands for cognitive control over task switching (learned switch readiness) are generalizable to a small degree, as associations that impact proficiency of forced switches can also influence preference for voluntary switches. However, learned switch readiness is task-specific, as indicated by the immediate elimination of switching proficiency when new tasks are used. Such task-specificity is consistent with the CA mechanism of learned switch readiness, and verifying this mechanism helps to clarify persisting confusion in the literature about what precisely is meant by cognitive flexibility (Ionescu, 2012). Moreover, the current findings help explain the lack of success for computerized brain training interventions aimed at improving flexibility (Buitenweg et al., 2017; Park & Bischof, 2013; Simons et al., 2016), as CA is likely the foundation for commercial flexibility training programs like Lumosity's *Brain Shift* (Bainbridge & Mayer, 2018). If a large degree of generalization is truly desired, future training programs will need to shift away from the CA-based mechanism and target non-CA based ones, possibly including WMU thresholds. On the other hand, returning to the driving example, while the human brain can learn to associate a particular busy intersection with an increased need to switch tasks while driving, walking through that same intersection may not always require the same amount of flexibility. Thus, task specificity of learned flexibility is perhaps beneficial, as overgeneralization of cognitive flexibility could be just as detrimental as deficiencies in flexibility.

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

Allport, D. A., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 421-452). Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/1478.003.0025

Arbuthnott, K., & Frank, J. (2000). Executive control in set switching: Residual switch cost and task-set inhibition. *Canadian Journal of Experimental Psychology, 54*(1), 33-41. https://doi.org/10.1037/h0087328

Arrington, C. M., & Logan, G. D. (2004). The cost of a voluntary task switch. *Psychological Science*, *15*(9), 610-615. https://doi.org/10.1111/j.0956-7976.2004.00728.x

Arrington, C. M., & Logan, G. D. (2005). Voluntary task switching: Chasing the elusive homunculus. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(4), 683-702. https://doi.org/10.1037/0278-7393.31.4.683

Bainbridge, K., & Mayer, R. E. (2018). Shining the light of research on Lumosity. *Journal of Cognitive Enhancement*, *2*(1), 43-62. https://doi.org/10.1007/s41465-017-0040-5

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624-652. https://doi.org/10.1037/0033-295X.108.3.624

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

Braem, S., Bugg, J. M., Schmidt, J. R., Crump, M. J., Weissman, D. H., Notebaert, W., & Egner, T. (2019). Measuring adaptive control in conflict tasks. *Trends in Cognitive Sciences*, *23*(9), 769-783. https://doi.org/10.1016/j.tics.2019.07.002

Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. In A. R. A. Conway, C. Jarrold, M. J. Kane, & J. N. Towse (Eds.), *Variation in working memory* (pp. 76-106). Oxford, United Kingdom: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195168648.003.0004

Brosowsky, N. P., & Egner, T. (2021). Appealing to the cognitive miser: Using demand avoidance to modulate cognitive flexibility in cued and voluntary task switching. *Journal of Experimental Psychology: Human Perception and Performance*, *47*(10), 1329-1347. https://doi.org/10.1037/xhp0000942

Bugg, J. M., & Crump, M. J. (2012). In support of a distinction between voluntary and stimulus-driven control: A review of the literature on proportion congruent effects. *Frontiers in Psychology*, *3*, 1-16. https://doi.org/10.3389/fpsyg.2012.00367

Bugg, J. M., Jacoby, L. L., & Toth, J. P. (2008). Multiple levels of control in the Stroop task. *Memory & Cognition*, *36*(8), 1484-1494. https://doi.org/10.3758/MC.36.8.1484

Buitenweg, J. I. V., van de Ven, R. M., Prinssen, S., Murre, J. M. J., & Ridderinkhof, K. R. (2017). Cognitive flexibility training: A large-scale multimodal adaptive active-control intervention study in healthy older adults. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00529

Chiew, K. S., & Braver, T. S. (2014). Dissociable influences of reward motivation and positive emotion on cognitive control. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(2), 509-529. https://doi.org/10.3758/s13415-014-0280-0

Chiu, Y. C., & Egner, T. (2019). Cortical and subcortical contributions to context-control learning. *Neuroscience & Biobehavioral Reviews*, *99*, 33-41. https://doi.org/10.1016/j.neubiorev.2019.01.019

Chiu, Y. C., Fröber, K., & Egner, T. (2020). Item-specific priming of voluntary task switches. *Journal of Experimental Psychology: Human Perception and Performance*, *46*(4), 434-441. https://doi.org/10.1037/xhp0000725

Chiu, Y.-C., & Egner, T. (2017). Cueing cognitive flexibility: Item-specific learning of switch readiness. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(12), 1950-1960. https://doi.org/10.1037/xhp0000420

Crump, M. J. C., & Logan, G. D. (2010). Contextual control over task-set retrieval. *Attention, Perception, & Psychophysics.* 72, 2047–2053. https://doi.org/10.3758/BF03196681

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*(1), 135-168. https://doi.org/10.1146/annurev-psych-113011-143750

Dreisbach, G. (2006). How positive affect modulates cognitive control: The costs and benefits of reduced maintenance capability. *Brain and Cognition*, *60*(1), 11-19. https://doi.org/10.1016/j.bandc.2005.08.003

Dreisbach, G., & Fröber, K. (2019). On how to be flexible (or not): Modulation of the stability-flexibility balance. *Current Directions in Psychological Science*, *28*(1), 3-9. https://doi.org/10.1177/0963721418800030

Dreisbach, G., & Goschke, T. (2004). How positive affect modulates cognitive control: Reduced perseveration at the cost of increased distractibility. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 343-353. https://doi.org/10.1037/0278-7393.30.2.343

Dreisbach, G., & Haider, H. (2006). Preparatory adjustment of cognitive control in the task switching paradigm. *Psychonomic Bulletin & Review*, *13*(2), 334-338. https://doi.org/10.3758/bf03193853

Fischer, R., & Hommel, B. (2012). Deep thinking increases task-set shielding and reduces shifting flexibility in dual-task performance. *Cognition*, *123*(2), 303-307. https://doi.org/10.1016/j.cognition.2011.11.015

Fröber, K., Pfister, R., & Dreisbach, G. (2019). Increasing reward prospect promotes cognitive flexibility: Direct evidence from voluntary task switching with double registration. *Quarterly Journal of Experimental Psychology*, *72*(8), 1926-1944. https://doi.org/10.1177/1747021818819449

Geurts, H. M., Begeer, S., & Stockmann, L. (2009). Brief report: Inhibitory control of socially relevant stimuli in children with high functioning autism. *Journal of Autism and Developmental Disorders*, *39*(11), 1603-1607. https://doi.org/10.1007/s10803-009-0786-4

Goschke, T. (2013). Volition in action: Intentions, control dilemmas and the dynamic regulation of intentional control. In W. Prinz, A. Beisert & A. Herwig (Eds.), *Action science: Foundations of an emerging discipline* (pp. 411-434). Cambridge, MA: MIT Press.

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

Hommel, B. (2015). Between persistence and flexibility: The yin and yang of action control. In A. Elliot (Ed.), *Advances in motivation science* (Vol. 2, pp. 33-67). New York, NY: Elsevier. https://doi.org/10.1016/bs.adms.2015.04.003

Ionescu, T. (2012). Exploring the nature of cognitive flexibility. *New Ideas in Psychology*, *30*(2), 190-200. https://doi.org/10.1016/j.newideapsych.2011.11.001

Isen, A. M. (2001). An influence of positive affect on decision making in complex situations: Theoretical issues with practical implications. *Journal of Consumer Psychology*, *11*(2), 75-85. https://doi.org/10.1207/S15327663JCP1102_01

Janczyk, J. M. (2016). Sequential modulation of backward crosstalk and task-shielding in dual-tasking. *Journal of Experimental Psychology: Human Perception and Performance, 42*(5), 631-647. https://doi.org/10.1037/xhp0000170

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, United Kingdom: Oxford University Press.

Jersild, A. T. (1927). Mental set and shift. *Archives of Psychology, 89*, 5-81.

Jiang, J., Beck, J., Heller, K., & Egner, T. (2015). An insula-frontostriatal network mediates flexible cognitive control by adaptively predicting changing control demands. *Nature Communications*, *6*, 8165. https://doi.org/10.1038/ncomms9165

Jiang, J., Heller, K., & Egner, T. (2014). Bayesian modeling of flexible cognitive control. *Neuroscience & Biobehavioral Reviews*, *46*, 30-43. https://doi.org/10.1016/j.neubiorev.2014.06.001

Kang, M. S., & Chiu, Y.-C. (2021). Proactive and reactive metacontrol in task switching. *Memory and Cognition*. https://doi.org/10.3758/s13421-021-01189-8

Koch, I., Gade, M., Schuch, S., & Philipp, A. M. (2010). The role of inhibition in task switching: A review. *Psychonomic Bulletin & Review, 17*, 1-14. https://doi.org/10.3758/PBR.17.1.1

Leboe, J. P., Wong, J., Crump, M., & Stobbe, K. (2008). Probe-specific proportion task repetition effects on switching costs. *Perception & Psychophysics*, *70*(6), 935-945. https://doi.org/10.3758/pp.70.6.935

Lien, M. C., & Ruthruff, E. (2008). Inhibition of task set: Converging evidence from task choice in the voluntary task-switching paradigm. *Psychonomic Bulletin & Review*, 15(6), 1111-1116. https://doi.org/10.3758/PBR.15.6.1111

Locke, H. S., & Braver, T. S. (2008). Motivational influences on cognitive control: Behavior, brain activation, and individual differences. *Cognitive, Affective, & Behavioral Neuroscience*, *8*(1), 99-112. https://doi.org/10.3758/CABN.8.1.99

Mayr, U., & Keele, S. W. (2000). Changing internal constraints on action: The role of backward inhibition. *Journal of Experimental Psychology: General*, *129*(1), 4-26. https://doi.org/10.1037/0096-3445.129.1.4

Mayr, U., & Kliegl, R. (2003). Differential effects of cue changes and task changes on task-set selection costs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(3), 362-372. https://doi.org/10.1037/0278-7393.29.3.362

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., Diamond, G. M., Toder, D., & Nemets, B. (2011). Cognitive rigidity in unipolar depression and obsessive compulsive disorder: Examination of task switching, Stroop, working memory updating and post-conflict adaptation. *Psychiatry Research*, *185*(1-2), 149-156. https://doi.org/10.1016/j.psychres.2010.04.044

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*(3), 134-140. http://dx.doi.org/10.1016/S1364-6613(03)00028-7

Monsell, S., & Mizon, G. A. (2006). Can the task-cuing paradigm measure an endogenous task-set reconfiguration process? *Journal of Experimental Psychology: Human Perception and Performance*, *32*(3), 493-516. https://doi.org/10.1037/0096-1523.32.3.493

Moreno-Martínez, F. J., & Montoro, P. R. (2012). An ecological alternative to Snodgrass & Vanderwart: 360 high quality colour images with norms for seven psycholinguistic variables. *PloS One*, *7*(5), E37527. https://doi.org/10.1371/journal.pone.0037527

Nack, C. A. (2022). *Item-specific switch conditioning to reduce switch costs* [Unpublished manuscript]. Department of Psychological Sciences, Purdue University.

Park, D. C., & Bischof, G. N. (2013). The aging mind: Neuroplasticity in response to cognitive training. *Dialogues in Clinical Neuroscience*, *15*(1), 109-119. 10.31887/DCNS.2013.15.1/dpark

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195-203. https://doi.org/10.3758/s13428-018-01193-y

Philipp, A. M., Jolicoeur, P., Falkenstein, M., & Koch, I. (2007). Response selection and response execution in task switching: Evidence from a go-signal paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(6), 1062-1075. https://doi.org/10.1037/0278-7393.33.6.1062

Plessow, F., Schade, S., Kirschbaum, C., & Fischer, R. (2012). Better not to deal with two tasks at the same time when stressed? Acute psychosocial stress reduces task shielding in dual-task performance. *Cognitive, Affective, & Behavioral Neuroscience*, *12*(3), 557-570. https://doi.org/10.3758/s13415-012-0098-6

Ravizza, S. M., & Carter, C. S. (2008). Shifting set about task switching: Behavioral and neural evidence for distinct forms of cognitive flexibility. *Neuropsychologia*, *46*(12), 2924-2935. https://doi.org/10.1016/j.neuropsychologia.2008.06.006

Robert, C. P., Chopin, N., & Rousseau, J. (2009). Harold Jeffreys's theory of probability revisited. *Statistical Science*, *24*(2), 141-172. https://doi.org/10.1214/09-sts284

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*(2), 207-231. http://dx.doi.org/10.1037/0096-3445.124.2.207

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356-374. https://doi.org/10.1016/j.jmp.2012.08.001

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225-237. https://doi.org/10.3758/PBR.16.2.225

Sabah, K., Dolk, T., Meiran, N., & Dreisbach, G. (2019). When less is more: Costs and benefits of varied vs. fixed content and structure in short-term task switching training. *Psychological Research*, *83*(7), 1531-1542. https://doi.org/10.1007/s00426-018-1006-7

Schneider, D. W. (2007). Task-set inhibition in chunked task sequences. *Psychonomic Bulletin & Review*, *14*(5), 970-976. https://doi.org/10.1037/e527342012-355

Schneider, D. W. (2015). Isolating a mediated route for response congruency effects in task switching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(1), 235-245. http://dx.doi.org/10.1037/xlm0000049

Schneider, D. W. (2018). Categorization difficulty modulates the mediated route for response selection in task switching. *Psychonomic Bulletin & Review*, *25*(5), 1958-1967. https://doi.org/10.3758/s13423-017-1416-3

Schneider, D. W., & Logan, G. D. (2006). Priming cue encoding by manipulating transition frequency in explicitly cued task switching. *Psychonomic Bulletin & Review*, *13*, 145-151. https://doi.org/10.3758/BF03193826

Schneider, D. W., & Logan, G. D. (2009). Task switching. *Encyclopedia of Neuroscience, 7*(3), 869-874. https://doi.org/10.1016/B978-008045046-9.00426-5

Shiffrin, R. D., & Schneider, W. (1977). Controlled and automatic human information processing, II: Perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*, 127-190. https://doi.org/10.1037/0033-295X.84.2.127

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. (2016). Do “brain-training” programs work? *Psychological Science in the Public Interest*, *17*(3), 103-186. https://doi.org/10.1177/1529100616661983

Siqi-Liu, A., & Egner, T. (2020). Contextual adaptation of cognitive flexibility is driven by task- and item-level learning. *Cognitive, Affective, & Behavioral Neuroscience*, *20*, 757-782. https://doi.org/10.3758/s13415-020-00801-9

Spector, A., & Biederman, I. (1976). Mental set and mental shift revisited. *American Journal of Psychology*, *89*(4), 669-679. https://doi.org/10.2307/1421465

van Wouwe, N. C., Band, G. P. H., & Ridderinkhof, K. R. (2011). Positive affect modulates flexibility and evaluative control. *Journal of Cognitive Neuroscience, 23*(3), 524-539. https://doi.org/10.1162/jocn.2009.21380

Wylie, G., & Allport, A. (2000). Task switching and the measurement of “switch costs”. *Psychological Research*, *63*, 212-233. https://doi.org/10.1007/s004269900003

Zwosta, K., Hommel, B., Goschke, T., & Fischer, R. (2013). Mood states determine the degree of task shielding in dual-task performance. *Cognition & Emotion*, *27*(6), 1142-1152. https://doi.org/10.1080/02699931.2013.772047