



Published in final edited form as:

*J Exp Psychol Hum Percept Perform.* 2023 August ; 49(8): 1111–1122. doi:10.1037/xhp0001138.

## Task sets define boundaries of learned cognitive flexibility in list-wide proportion switch manipulations

Audrey Siqu-Liu<sup>1,2</sup>, Tobias Egner<sup>1,2</sup>

<sup>1</sup>Department of Psychology and Neuroscience, Duke University, Durham, NC 27708

<sup>2</sup>Center for Cognitive Neuroscience, Duke University, Durham, NC 27705

### Abstract

Different contexts in daily life often require varying levels of cognitive flexibility. Previous research has shown that people adapt their level of flexibility to match changing contextual demands for task switching in cued-switching paradigms that vary the proportion of switch trials within lists of trials. Specifically, the behavioral costs of switching as opposed to repeating tasks scale inversely with the proportion of switches – a finding referred to as the list-wide proportion switch (LWPS) effect. Previous research found that flexibility adaptations transferred across stimuli, but were specifically tied to task-sets, rather than block-wide changes in flexibility state. In the current study, we conducted additional tests of the hypothesis that flexibility learning is task-specific in the LWPS paradigm. In Experiments 1 and 2, we used trial-unique stimuli and unbiased task cues to control for associative learning tied to stimulus or cue features. Experiment 3 further tested whether task-specific learning occurred even for tasks performed on integrated features of the same stimuli. Across these three experiments, we found robust task-specific flexibility learning, which transferred across novel stimuli and unbiased cues and occurred regardless of stimulus feature overlap between tasks.

### Keywords

cognitive control; flexibility adaptation; task set

### Introduction

Different activities in daily life often require different levels of cognitive flexibility, or readiness to switch between tasks. For example, cooking may require rapid switching between several different tasks like monitoring your pasta water and chopping up tomatoes for a sauce. In other situations, it may be more important to focus on the current task and ignore potential distractions in the environment, such as when a student is reviewing for an exam in a loud college dormitory room. Previous research has shown that people can adapt their level of flexibility (or switch-readiness) to match changing contextual demands for task

---

Correspondence concerning this article should be addressed to Audrey Siqu-Liu, Duke University, Center for Cognitive Neuroscience, Levine Science Research Center Box 90999, Durham, NC 27708-0999, U.S.A. [audrey.liu@duke.edu](mailto:audrey.liu@duke.edu).

Experimental code, data, and analyses are openly available at the project's Open Science Framework page (<https://osf.io/ueqa3/>). This study was funded by NIMH R01 MH116967. We have no conflicts of interest to disclose. This study was not preregistered.

switching (Bonnin et al., 2011; Dreisbach & Haider, 2006; Dreisbach et al., 2002; Duthoo et al., 2012; Mayr, 2006; Monsell & Mizon, 2006; Siqi-Liu & Egner, 2020), and a range of psychiatric conditions are associated with an impaired regulation of cognitive flexibility (Uddin, 2021). To improve our understanding of how people regulate their switch-readiness, the current study investigated the boundary conditions of learned flexibility.

Prior studies of adaptation in flexibility used cued task switching paradigms and manipulated the frequency of task switches across blocks or lists of trials to create contexts with a “high” proportion of task switches (high PS contexts), demanding a high level of cognitive flexibility, and “low” PS contexts, where demands on cognitive flexibility were low. The canonical finding is that the behavioral cost of switching – slower and more error-prone responding on task switch than repeat trials - scales inversely with PS: participants incur smaller switch costs, indicating greater flexibility, in high than low PS contexts (Bonnin et al., 2011; Dreisbach & Haider, 2006; Dreisbach et al., 2002; Duthoo et al., 2012; Mayr, 2006; Monsell & Mizon, 2006; Siqi-Liu & Egner, 2020). This finding has been referred to as the list-wide proportion switch (LWPS) effect (Siqi-Liu & Egner, 2020).

An intuitive interpretation of the LWPS effect is that participants enter a sustained state of increased/decreased flexibility which lasts throughout high/low PS contexts. However, high PS blocks usually contain tasks and stimuli that are also more frequently presented as switch trials, and vice versa for low PS blocks and repeat trials. Without de-confounding switch-associations at the task and stimulus-level, the presumed “block-wide” changes in flexibility could also be driven by participants learning to perform more flexibly in response to *specific* tasks or stimuli that they have experienced more often as switch trials, as opposed to them being in a tonically more flexible cognitive state.

To investigate whether flexibility learning is tied to blocks, tasks, or stimuli, Siqi-Liu & Egner (2020) introduced cued-task switching paradigms that included transfer task and stimulus sets, which were unbiased in terms of switch/repeat- associations (i.e., they were presented equally often as task-switch and repeat trials). These unbiased sets were embedded within overall switch/repeat biased LWPS contexts to probe whether the inverse relationship between switch cost and PS context could be identified not only in the biased task/stimulus sets, but also in the unbiased transfer sets.

Specifically, Experiment 4 of Siqi-Liu & Egner (2020) controlled for stimulus bias in the following manner. Depending on the task cue, participants performed either a digit task, where they categorized single digits (from 2–9) as odd or even, or a letter task, where they categorized letters (A, E, I, U, G, K, M, or R) as vowels or consonants. On each trial, one digit and one letter were selected for stimulus presentation. To create an unbiased stimulus set, a subset of the digits (e.g., 2, 6, 9, and 3) and letters (e.g., A, I, G, and M) was randomly pre-selected to be presented equally often on switch and repeat trials in both the high and low PS context blocks. The remaining digits (i.e., 4, 8, 5, and 7) and letters (E, U, K, and R) served as biased stimuli and occurred more often as switch trials in high PS blocks and more often as repeat trials in low PS blocks, thus driving the overall PS context manipulation at the list-wide level.

Experiments 3a and 3b of Siqi-Liu & Egner (2020) controlled for task- instead of stimulus-bias in a similar manner. The same letter and digit tasks were used here, but a color categorization task, where participants responded whether a colored frame around the letter and digit stimuli was a warm or cold color, was added. One task was randomly pre-selected to be the unbiased task (e.g., letter categorization). The unbiased task was then presented equally often as switch and repeat trials across the different PS contexts, while the other two tasks (e.g., digit categorization and color categorization) occurred more often as switch trials in high PS blocks and more often as repeat trials in low PS blocks, thus driving the overall list-wide PS context.

Given these results, Siqi-Liu & Egner (2020) concluded that learned switch-readiness in the LWPS paradigm appears to be specifically tied to task sets rather than to stimuli-specific associations, or to block-level sustained state changes (see also Siqi-Liu, Egner, & Woldorff, 2022). In other words, people seem to associate different levels of flexibility with different task sets depending on how often they have had to switch to that task. However, that study left open several possible alternative explanations. In the current series of experiments, we conducted additional tests of the hypothesis that flexibility learning is task-specific in the LWPS paradigms.

First, to further determine the extent that flexibility learning is agnostic to stimulus-level associations, in Experiment 1, we employed a task design with trial-unique stimuli as a novel and straightforward way to test for flexibility adaptations in settings where forming stimulus-switch associations is impossible. The same cued task-switching paradigm with a list-wide proportion switch manipulation was implemented, but completely novel stimuli were presented on each trial. Given prior evidence from Siqi-Liu & Egner (Exp 4, 2020), we hypothesized that we would replicate the LWPS effect in the absence of any stimulus repetitions altogether.

Second, though Siqi-Liu & Egner (2020, Exp 3a-b) suggested that flexibility learning is task-specific, it remains unclear whether people's switch-readiness in this study was linked to the physical task cues, which were also biased, or to the abstract task-set representations, i.e., the rules that define stimulus-response associations (Kiesel et al., 2010; Monsell, 2003; Rogers & Monsell, 1995). In Experiment 2, we asked participants to switch between two biased tasks that were each associated with two cues. One of the cues for each task was biased, i.e., it occurred more frequently on switch trials in high PS contexts and more frequently on repeat trials in low PS contexts, and the other cue was presented equally often on switch and repeat trials, regardless of PS context. Assuming that flexibility learning is tied to the actual task sets rather than the task cues, we expected to see switch cost adjustments to PS for both biased and unbiased cues.

Third, Siqi-Liu & Egner (2020, Exp 3a-b) investigated transfer of flexibility learning using tasks performed on distinct and spatially separated stimuli or stimulus features. That is, the letter, digit, and color categorization tasks used in that paradigm involved different stimuli (letters, digits, and colored frames, respectively) that appeared in distinct, non-overlapping locations on screen. Though the stimuli for all three tasks were presented on every trial, regardless of the cued task, they were not overlaid or integrated features of one visual

object (e.g., such as age, gender, and emotion dimensions of face stimuli). Given prior evidence of control transfer based on shared stimulus features (Bustamente et al., 2021), it is theoretically possible that flexibility could transfer across tasks performed on integrated feature dimensions of the same stimulus object. We tested this idea in Experiment 3 by instructing participants to categorize faces along three different feature dimensions, such that the transfer task was performed on the same, integrated stimuli as the biased tasks (using age/gender/emotion dimensions of face stimuli).

By testing whether flexibility settings generalize to unbiased tasks that appear alongside biased tasks in temporal contexts that pose varying switch demands, the current series of experiments seek to enhance our understanding of the environmental and task conditions that encourage task-specific flexibility learning over global flexibility adaptations.

## Experiment 1

In Siqi-Liu & Egner (2020) as well as in the LWPS literature at large (Bonnin et al., 2011; Dreisbach & Haider, 2006; Dreisbach et al., 2002; Duthoo et al., 2012; Mayr, 2006; Monsell & Mizon, 2006; Siqi-Liu & Egner, 2020), cued-switching paradigms always involve recurring stimuli, at least some of which were associated with a switch- or repeat-bias. It is therefore unknown whether an LWPS effect can be observed in the complete absence of any stimulus repetitions.

In the current experiment, we used trial-unique stimuli as a novel way to control for item-based learning effects in list-wide PS paradigm. Since each stimulus only appeared once, there was no opportunity for stimulus-switch associations to form. We used the same basic design as Siqi-Liu & Egner (2020, Exp 2) and manipulated LWPS context while participants switched between two object categorization tasks. Each trial consisted of a unique object stimulus. We hypothesized that participants would nevertheless exhibit lower switch costs in the high compared to low PS contexts, providing decisive evidence that flexibility adjustments do not depend on any associations between specific stimuli and control states.

## Methods

**Participants**—We conducted a power analysis (G\*Power 3.1; Faul et al., 2009) using the effect size for the LWPS effect in response times (RT) found in Siqi-Liu & Egner (2020, Exp 2, partial eta squared ( $\eta_p^2$ ) = 0.28), which employed a similar cued-switching paradigm involving two tasks. This revealed that a sample size of 23 participants would be necessary for 80% statistical power to detect the effect at  $p < 0.05$ . We therefore targeted a sample size of 30 participants from the Duke University Department of Psychology and Neuroscience Subject Pool who surpassed an accuracy inclusion criterion of  $> 65\%$ . There were 20 female participants and 10 male participants. Average age was 19.2 with a standard deviation ( $SD$ ) of 1.03 years. All participants gave informed consent and received payment or course credit in accordance with a protocol approved by the Duke University Institutional Review Board.

**Stimuli**—Experimental stimuli consisted of 427 pictures of objects gathered using Google Image Search (taken from Wen & Egner, 2022, Exp 3 & 4). Objects were either manmade

(e.g., a car) or natural (e.g., a tree) and either smaller than a shoe box (e.g., a key) or larger than a shoebox (e.g., a house), resulting in four stimulus categories. There were 107 each of small-manmade, large-manmade, large-natural, and 106 small-natural objects.

**Procedures**—Each trial began with a central fixation cross displayed in black on a grey background for 300 ms. The fixation cross then turned green or blue to cue the upcoming task. If the fixation turned green, participants had to perform the *origin* task and categorize the upcoming object as natural or manmade. If the fixation turned blue, participants had to perform the *size* task and categorize the object as larger or smaller than a shoebox. After 160 ms, the object stimulus was displayed centrally behind the fixation cross. The object was randomly selected on each trial without replacement from the 427 stimuli. The fixation and stimulus stayed on screen for 1200 ms, during which participants used left or right arrow key presses to indicate their response to the task. Responses during the stimulus presentation duration were logged, and feedback (“correct”, “incorrect”) was presented after stimulus offset for 500 ms.

These trial timing parameters were comparable to Experiment 3a of Siqi-Liu & Egner (2020), which included a blank screen of 1010 ms, followed by a 450 ms fixation cross display, and a 160 ms cue display. In that experiment, the cue was removed from screen 40 ms before stimulus onset. Task stimuli were then displayed for 1300 ms and follow by either a 2000 ms error feedback display or a 500 ms blank screen if participants responded correctly. In contrast, the current experiment had a shorter pre-cue (blank + fixation) display and stimulus display and presented both correct and incorrect feedback after stimulus offset for 500 ms. These adjustments to the trial timing were made to simplify the task design and account for decreased task difficulty of the current picture classification task compared to the letter/number/color classification task in Siqi-Liu & Egner (2020, Exp 3).”

Participants completed 1 practice block and 6 experimental blocks of 61 trials each (427 trials total). The practice block consisted of 50% switch trials and 50% repeat trials. Half of the experimental blocks had a low (30%) proportion of switches (PS) and half had a high (70%) PS. Participants were randomly assigned to complete all low or all high PS blocks first. A reminder of the correct response mappings was displayed between the blocks. Both tasks were presented an equal number of times in each block. All trial sequences were pseudo-randomly generated by a JavaScript algorithm.

**Transparency and Openness**—We report how we determined our sample size, data exclusions, manipulations, and experimental measures, and follow JARS (Kazak, 2018). Data, task and analysis code for all experiments are available at the project’s Open Science Framework page (<https://osf.io/ueqa3/>). Data were analyzed using R, version 3.6.1 (R Core Team, 2019). The study’s design and analysis were not preregistered.

## Results

To assess performance accuracy, we analyzed data from all trials after excluding practice blocks and the first trial of each experimental block. For RT analyses, we additionally excluded incorrect trials and post-error trials. After applying these exclusion criteria, trials with response times outside 1.5 times the interquartile range were filtered out.

We conducted a 2 (PS: low v. high)  $\times$  2 (trial type: repeat v. switch) mixed ANOVA on RTs and accuracy. The data are summarized in Figure 2. Responses were slower ( $F(1,29) = 73.93$ ,  $p < .001$ ,  $\eta^2 = .72$ ) and more error-prone ( $F(1,29) = 62.82$ ,  $p < .001$ ,  $\eta^2 = .68$ ) on switch ( $M_{\text{rt}} = 729.9.1$  ms;  $M_{\text{error}} = 0.24$ ) compared to repeat trials ( $M_{\text{rt}} = 679.2$  ms;  $M_{\text{error}} = 0.15$ ). Consistent with previous reports of the LWPS effect, trial type significantly interacted with PS in response times ( $F(1,29) = 7.72$ ,  $p = .009$ ,  $\eta^2 = .21$ ) but not error rates. This interaction reflected smaller RT switch costs in high PS blocks ( $M_{\text{switchcost}} = 40.0$  ms) compared to low PS blocks ( $M_{\text{switchcost}} = 68.3$  ms), representing the LWPS effect.

## Discussion

To probe whether flexibility adjustments to varying PS could be obtained in the absence of any possible stimulus-task or stimulus-switch associations, Experiment 1 consisted of a standard LWPS design but with trial-unique stimuli. The fact that we nevertheless obtained a typical LWPS effect provides further evidence that the contextual adaptation of switch-readiness can occur independently of any stimulus-level learning of control demands. Rather, flexibility learning seemed to transfer to completely novel stimuli upon first exposure.

These findings are consistent with results from the conflict control literature with experiments that utilize conceptually similar designs of manipulating the proportion of high conflict trials in a list to measure conflict adaptation effects (“list-wide proportion congruency effects”; for review see (Bugg & Crump, 2012). Using nonrepeating stimuli in a picture-word interference paradigm, Spinelli et al. (2019) observed proportion congruent effects in the absence of stimulus-level associations, similar to our observations of flexibility adaptations with trial-unique task-switching stimuli. In sum, the Experiment 1 results indicate that list-wide adaptation of flexibility does not rely on any kind of stimulus-level contingency learning.

## Experiment 2

Just as target stimuli could be associated with switch- or repeat biases, task-cues are stimulus features of a trial that could also be targets for flexibility learning. For instance, Bejjani et al. (2018) found that task-irrelevant scenes that were either more frequently associated with congruent or incongruent trials cued control adaptations to upcoming Stroop-task stimuli, such that smaller conflict was found in Stroop trials preceded by scenes that were mostly paired with incongruent stimuli. Since trial features preceding a target stimulus can thus cue control processes, it is possible that previous observations of task-specific flexibility learning in the LWPS paradigm could be the result of learned flexibility associations with physical features of the task-cue, rather than the abstract task-set (i.e., the set of rules that define stimulus-response associations). Although Siqi-Liu & Egner (2020, Exp 3a-b) included separate biased and unbiased tasks, within the biased task-sets, the relevant task cues were also more frequently associated with task switches in high PS contexts and with task repeats in the low PS contexts. In other words, participants could have also retrieved specific flexibility states in response to the physical task cues (e.g., the word



“letter”), rather than retrieving them in response to the meaning of the cue, or the cued task set.

To determine whether flexibility learning depends on switch/repeat-associations with specific task cues, in the current experiment, we isolated the effects of cue-bias and investigated whether Siqi-Liu & Egner’s (2020, Exp 3) pattern of results could be replicated with biased and unbiased task cues rather than biased and unbiased tasks. Based on the assumption that learned switch-readiness is linked to task sets rather than their (arbitrary) cues, we predicted that flexibility state *would* transfer to unbiased task cues, even though it did not transfer to unbiased task sets (Siqi-Liu & Egner, 2020). This would indicate that flexibility learning not tied the specific visual features that cue the task and provide further evidence that flexibility learning is specifically tied to the task rules themselves.

## Methods

**Participants**—We conducted a power analysis (G\*Power 3.1; Faul et al., 2009) based on the effect size for the three-way interaction between trial type  $\times$  PS  $\times$  task bias in RT data found in Siqi-Liu & Egner (2020, Exp 3a,  $\eta_p^2 = 0.09$ ). This revealed that a sample size of 82 participants would be necessary for 80% statistical power to detect the effect at  $p < 0.05$ . Based on these power analysis results, our final sample for data analysis consisted of a new group of 83 participants from the Duke University Department of Psychology and Neuroscience Subject Pool who surpassed an accuracy inclusion criterion of  $> 65\%$ . There were 53 female participants, 27 male participants, and 3 participants who did not provide gender information. Average age of participants was 18.9 with an *SD* of 0.98 years. All participants gave informed consent and received payment or course credit in accordance with a protocol approved by the Duke University Institutional Review Board.

**Stimuli**—Participants switched between letter (“is the letter a consonant or vowel?”) and digit (is the number odd or even?) classification tasks. The letter task was cued by the words “letter” and “alphabet” and the digit task was cued by the words “digit” and “number.” For each participant, one of the cues for each task was randomly selected to be a biased cue and one was selected as an unbiased transfer cue.

On each trial, one of eight letters (A, E, I, U, G, K, M, R) and one of eight digits (2, 3, 4, 5, 6, 7, 8, 9) were randomly selected and presented simultaneously to the left and right of a fixation cross. The location of letter and digit stimuli was randomized on a trial-by-trial basis.

**Procedures**—All trial timing parameters except the error feedback duration were identical to Siqi-Liu & Egner (2020, Exp 3a). Each trial began with a blank screen lasting 1010 ms, followed by a fixation cross displayed for 450 ms. The cue was then displayed for 160 ms, followed by another blank screen lasting 40 ms, and a stimulus display of 1300 ms during which participants used a ‘d’ or ‘k’ key press to indicate whether the stimulus was a consonant/vowel or odd/even based on instructions received at the beginning of the task. Responses during the stimulus presentation duration were logged. In Siqi-Liu & Egner (2020, Exp 3a), participants viewed either a 200 ms error feedback display for incorrect responses, or a 500 ms blank screen for correct responses. Here, to simplify the experimental

design, textual feedback (“correct”, “incorrect”) was presented after stimulus offset for 500 ms on each trial.

All participants completed 1 practice block and 14 experimental blocks of 33 trials each. The practice block consisted of 50% switch trials and 50% repeat trials. Half of the experimental blocks had a low (30%) proportion of switches (PS) and half had a high (70%) PS. Participants were randomly assigned to complete either all the low or all the high PS blocks first. In between blocks, participants could take a short self-paced break, during which a reminder of the response mappings was displayed.

In the low PS blocks, each task was presented 16 times total, 11 times as repeat trials and five times as switch trials. Of the 11 times each task appeared as a repeat trial, it was cued by the biased cue seven times and the unbiased cue four times. Out of the five times each task appeared as a switch trial, it was cued by the biased cue one time and the unbiased cue four times. Thus, the unbiased cue was associated with switch and repeat trials an equal number of times, while the biased cue appeared more often in association with repeat more than with switch trials in the low PS blocks. The same trial counts were used for the high PS blocks, only with the count for switch and repeat trials reversed. Thus, the unbiased cue was equally often associated with switch and repeat trials whereas the biased cue appeared more often in association with switch than with repeat trials in high PS blocks. These trial sequences were pseudo-randomly generated.

## Results

The same exclusion criteria for accuracy and RT analyses as Experiment 1 were applied. We conducted a 2 (PS: high v. low)  $\times$  2 (trial type: switch v. repeat)  $\times$  2 (cue bias: biased v. transfer) mixed ANOVA on response times and accuracy. The data are summarized in Figure 3.

Responses were slower ( $F(1,82) = 32.39, p < .001, \eta^2 = .28$ ) and more error-prone ( $F(1,82) = 49.09, p < .001, \eta^2 = .37$ ) on switch ( $M_{rt} = 812.1$  ms;  $M_{error} = 0.18$ ) compared to repeat trials ( $M_{rt} = 789.1$  ms;  $M_{error} = 0.13$ ). Trial type significantly interacted with PS in response times ( $F(1,82) = 7.69, p = .007, \eta^2 = .10$ ) but not error rates. Consistent with previous reports of the LWPS effect, this interaction reflected smaller RT switch costs when PS was high ( $M_{switchcost} = 11.5$ ) blocks compared to when PS was low ( $M_{switchcost} = 27.5$ ). Notably, there was no three-way interaction of trial type  $\times$  PS  $\times$  cue bias ( $F(1,82) = 0.04, p = .84, \eta^2 < .001$ ), indicating that the LWPS effect did not significantly differ between trials containing biased and unbiased cues.

## Discussion

In sum, the results from Experiment 2 replicated the LWPS effect and suggest that flexibility adjustments in the LWPS protocol were not moderated by switch-frequency learning tied to the task cue. Though flexibility adjustments did not transfer to the unbiased task set in Siqi-Liu & Egner (2020), we demonstrate here that they did transfer to unbiased task cues associated with a biased task set (see Figure 3). These results suggest that switch frequency learning is specifically tied to task-sets associated with frequent switching or repeating, as



speculated in Siqi-Liu & Egner (2020), rather than to the arbitrary physical stimulus features of the task cue.

### Experiment 3

The results of Experiment 2 indicate that flexibility learning is specifically tied to task-set representations rather than visual features of the task cues. Together with a lack of transfer to unbiased task sets found in Siqi-Liu & Egner (2020, Exp 3a-b), this provides further evidence for the conclusion that flexibility learning is task-specific in the list-wide PS protocol. However, the lack of transfer to the unbiased task set observed in Siqi-Liu & Egner (2020, Exp 3a-b) may have had to do with the fact that the letter, digit, and color tasks used in that study were each performed on rather distinct and spatially separated stimuli or stimulus features, which may have encouraged task-specific flexibility learning.

There is some prior evidence of control transfer based on shared stimulus features. Bustamante et al. (2021) demonstrated that transfer occurred when novel stimuli consisted of composite features that were previously associated with incentives for control exertion. Specifically, participants initially learned that one set of features in a color-word Stroop task (e.g., the color **blue** and the word RED) predicted that the color naming task would be rewarded, and that another set of features predicted that the word reading task would be rewarded. Then, the experimenters introduced new stimuli (e.g., **RED**) that combined features that were previously associated with reward for the color-naming task but together predicted that the word-reading task would be rewarded. They found that participants overexerted control (chose to complete color naming more) for new stimuli containing features that previously predicted high value of control. Thus, feature-control state associations transferred to new, integrated stimuli. Transfer of control states has also been demonstrated across stimuli linked through associative learning (Bejjani et al., 2018) and across linked spatial locations (Weidler & Bugg, 2016). Given the evidence for feature-based transfer in control learning, it is possible that learned flexibility demands could transfer across tasks when they are conducted on overlapping features of integrated stimuli.

On the other hand, when task sets involve distinct, non-overlapping stimuli, participants may be more likely to see switching between task sets as an event boundary. For instance, (Wang & Egner, 2022) found that switching tasks created event boundaries, as documented via participants' temporal order and distance memory judgements of trial-unique stimuli. In a list-wide PS paradigm, participants may thus exhibit less cross-task transfer of learned flexibility when they perceive task-sets as stronger event boundaries.

In the current experiment, we therefore investigated whether task sets would still form the boundary of flexibility learning if the tasks involved categorizing different integrated feature dimensions of the same stimuli. For this purpose, we compared flexibility transfer between two scenarios in a between-groups design. One group performed three face feature categorization tasks, such that transfer was assessed between two biased face categorization tasks to one unbiased face categorization task (high level of stimulus overlap). The other group performed two facial categorization tasks and one object categorization task, and

transfer was tested from the two biased face categorization tasks to the object categorization task (i.e., no stimulus overlap).

## Methods

**Participants**—Power analysis (G\*Power 3.1; Faul et al., 2009) based on effect sizes from Siqi-Liu & Egner (2020, Exp 3a,  $\eta_p^2 = 0.09$ ) found that a sample size of 82 was necessary to detect a three-way interaction between trial type  $\times$  PS  $\times$  task bias at  $p < 0.05$  with 80% power. As such, we aimed to include at least 82 participants in each group. Our final sample for data analysis consisted of a new group of 170 participants from Amazon Mechanical Turk who performed above the accuracy criteria of  $>65\%$ . The overlap and non-overlap groups each consisted of 85 randomly assigned participants. There were 77 female participants, 81 male participants, and 2 participants who responded with “Other.” Average age of participants was 39.9 with an *SD* of 12.19 years. All participants gave informed consent and received payment or course credit in accordance with a protocol approved by the Duke University Institutional Review Board.

**Stimuli**—For the overlap group, stimuli consisted of 104 pictures of faces of unique identities (Ebner et al, 2010). Faces differed along three dimensions: age (younger/older than 30), gender (male/female), and emotion (happy/sad).

For the non-overlap group, stimuli consisted of 80 pictures of faces of unique identities (Ebner et al., 2010) and 40 pictures of objects that were either manmade (e.g., house) or natural (e.g., tree) (Wen & Egner, 2022).

**Procedures**—For each participant, two tasks were designated as biased tasks and one as the unbiased transfer task. In the overlap group, the two biased tasks and transfer task were selected from three face categorization tasks. The biased tasks were the “gender” task, where participants decided whether the face stimulus was male or female and the “age” task, which involved categorizing the face as young (roughly under 30) or old (roughly over 30). The transfer task was the “emotion” task, which involved deciding whether the expression on the face shown is happy or sad. On each trial, one face stimulus was randomly selected with replacement from the set of 104 faces.

In the non-overlap group, the biased tasks were gender and age tasks, performed on pictures of faces. On each trial consisting of an age or gender task, a face stimulus was randomly selected with replacement from the set of 80 faces. The transfer task was performed on a separate stimulus set of pictures of objects and consisted of an “origin” task where participants responded whether each object stimulus was manmade or natural. On each trial of the origin task, an object stimulus was randomly selected with replacement from the set of 40 object stimuli.

Within each experimental block, all participants randomly switched between the three tasks based on a preceding experimental cue. Each trial began with a fixation cross, displayed for 300 ms, followed by a cue word (“gender”, “age”, “emotion” or “origin”) that preceded the stimulus for 200 ms and indicated which task the participant had to perform on the upcoming stimulus. The overlap group only received gender, age, and emotion cues while

the non-overlap group only received gender, age, and origin cues. The face/object stimulus was then displayed for 1200 ms, during which the task cue stayed on screen superimposed on top of the face/object. During the stimulus display, participant used a left or right arrow key press to indicate whether the stimulus was young/old, happy/sad, male/female, or manmade/natural based on instructions for response mappings received at the beginning of the task. After stimulus offset, feedback (“correct,” “incorrect,” or “too slow”) was displayed for 300 ms. These timing parameters were similar to those of Experiment 1, except for a slightly longer cue display (200 ms compared to 160 ms) and a slightly shorter feedback duration (300 ms compared to 500 ms).

All participants completed 1 practice block and 16 experimental blocks of 31 trials each. Stimuli that were used in the practice block did not appear in the experimental blocks. Half of the experimental blocks had a low (30%) PS trial sequence and half had a high (70%) PS rate. Participants were randomly assigned to complete either all the low or all the high PS blocks first. In between blocks, participants could take a short self-paced break, during which a reminder of the response mappings was displayed.

To achieve an overall switch: repeat trial ratio of 9:21 (30%) in the low PS blocks, the two biased tasks were each presented twice as switch trials and eight times as repeat trials, and the transfer task was presented five times as switch and five times as repeat trials. For a 21:9 switch: repeat trial ratio (70%) in the high PS blocks, the number of switch versus repeat trials was reversed for the biased tasks, whereas the unbiased task was still presented five times as switch and five times as repeat trials. Thus, the transfer task appeared equally often as switch and repeat trials in both the high and low PS blocks. These trial sequences were pseudo-randomly generated.

## Results

The same trial exclusion and filtering criteria were applied as in experiment 1. We conducted a 2 (group: overlap v. non-overlap)  $\times$  2 (PS: high v. low)  $\times$  2 (trial type: switch v. repeat)  $\times$  2 (task bias: biased v. transfer) mixed ANOVA on response times and accuracy. The data are summarized in Figure 4.

Responses were slower ( $F(1,168) = 9.38, p = .003, \eta^2 = 0.053$ ) and more error-prone ( $F(1,168) = 23.61, p < .001, \eta^2 = .12$ ) in the overlap group ( $M_{rt} = 822.8$  ms;  $M_{error} = 0.18$ ) compared to the non-overlap group ( $M_{rt} = 788.6$  ms;  $M_{error} = 0.15$ ). There were significant switch costs in response times ( $F(1,168) = 258.6, p < .001, \eta^2 = .61$ ) and error rates ( $F(1,168) = 294.2, p < .001, \eta^2 = .64$ ), wherein switch trials were associated with longer response times and higher error rates ( $M_{rt} = 824.9$  ms;  $M_{error} = 0.20$ ) compared to repeat trials ( $M_{rt} = 787.4$  ms;  $M_{error} = 0.13$ ). On average, transfer tasks were associated with faster RTs and lower error rates ( $M_{rt} = 769.6$  ms;  $M_{error} = 0.15$ ) than biased tasks ( $M_{rt} = 823.9$ ;  $M_{error} = 0.17$ ) (RT:  $F(1,168) = 127.2, p < .001, \eta^2 = .43$ ; Err:  $F(1,168) = 24.33, p < .001, \eta^2 = .13$ ). No other main effects reached significance.

We found an interaction effect of group  $\times$  task bias in both RTs ( $F(1,168) = 206.4, p < .001, \eta^2 = .55$ ) and error rates ( $F(1,168) = 119.6, p < .001, \eta^2 = .42$ ). In the overlap group, the transfer task was associated with slower RTs and higher error rates ( $M_{rt} = 836.9$  ms;  $M_{error}$

= 0.21) than the biased tasks ( $M_{rt} = 816.2$  ms;  $M_{error} = 0.16$ ). This relationship was reversed in the non-overlap group, where the transfer task was instead associated with faster RTs and lower error rates ( $M_{rt} = 711.9$  ms;  $M_{error} = 0.08$ ) compared to the biased tasks ( $M_{rt} = 831.8$  ms;  $M_{error} = 0.18$ ). This interaction effects may reflect differences in task difficulty: the origin task (which was the transfer task for the non-overlap group) appeared to be somewhat easier than the face emotion tasks (transfer task for the overlap group). The face tasks may be more difficult in part because of increased cross-task interference due to stimulus overlap between the different face tasks.

Switch costs were significantly higher in the overlap group ( $M_{rt} = 45.0$  ms;  $M_{error} = 0.08$ ) compared to the non-overlap group ( $M_{rt} = 30.9$  ms;  $M_{error} = 0.04$ ) in both RTs ( $F(1,168) = 8.9$ ,  $p = .003$ ,  $\eta^2 = .05$ ) and error rates ( $F(1,168) = 24.4$ ,  $p < .001$ ,  $\eta^2 = .13$ ). We also replicated the LWPS effect, finding a significant interaction effect of trial type  $\times$  PS in RTs ( $F(1,168) = 34.0$ ,  $p < .001$ ,  $\eta^2 = .17$ ) but not error rates. This was driven by smaller switch costs in the high PS condition ( $M_{rt} = 30.9$  ms) compared to the low PS condition ( $M_{rt} = 45.0$  ms).

Crucially, there was a three-way interaction of trial type  $\times$  proportion switch  $\times$  task bias in RTs ( $F(1,168) = 441.5$ ,  $p < .001$ ,  $\eta^2 = .12$ ) and error rates ( $F(1,168) = 10.5$ ,  $p = .001$ ,  $\eta^2 = .06$ ), indicating that the LWPS effect was modulated by task bias (see Figure 4). Furthermore, there was no four-way interaction with group in RTs or error rates, indicating that group did not significantly modulate the interaction between task bias and the LWPS effect, or that the LWPS effect was modulated by task bias regardless of whether the transfer task contained stimulus overlap with the biased tasks.

To further investigate how task bias affected LWPS effects in the two groups, we ran separate 2 (trial type)  $\times$  2 (PS) ANOVAs for each condition resulting from crossing task bias  $\times$  group.

In the overlap group, the trial type  $\times$  PS interaction was significant for the biased tasks in RTs ( $F(1,84) = 33.4$ ,  $p < .001$ ,  $\eta^2 = .28$ ) and error rates ( $F(1,84) = 13.4$ ,  $p < .001$ ,  $\eta^2 = .14$ ) but was nonsignificant for the transfer task in both RTs ( $F(1,84) = .13$ ,  $p = .72$ ,  $\eta^2 = .001$ ) and error rates ( $F(1,84) = 1.7$ ,  $p = .20$ ,  $\eta^2 = .02$ ).

Similarly, in the non-overlap group, the trial type  $\times$  PS interaction was significant for biased tasks in RTs ( $F(1,84) = 21.0$ ,  $p < .001$ ,  $\eta^2 = .20$ ), while the interaction representing the LWPS effect was nonsignificant for the transfer task ( $F(1,84) = 1.5$ ,  $p = .22$ ,  $\eta^2 = .02$ ). The LWPS effect did not reach significance in any of the error rate ANOVAs.

## Discussion

The results of Experiment 3 replicated the LWPS effect of reduced switch costs when participants switched more frequently; however, as in Siqi-Liu and Egner (2020), flexibility learning did not transfer to unbiased tasks presented within the same blocks, even when the transfer task shared overlapping stimulus features with the biased tasks. This indicates that task sets form the boundaries of switch-readiness learning in the LWPS paradigm, regardless of whether they are applied to distinct or identical stimuli.

Though there may have been some individual variance between participants' performance on the different tasks, this should not affect our interpretation of the LWPS effect, which is based on within-participant performance differences. However, one possible limitation of the current design is that different transfer tasks were used in the overlap and nonoverlap groups (the face emotion task and the object origin task, respectively), which makes it difficult to isolate performance differences due to stimulus overlap with the biased tasks from inherent differences between the two transfer tasks. As reported in the above results, the transfer task in the overlap group was associated with worse performance than the biased tasks while the transfer task in the nonoverlap group was associated with better performance than the same biased tasks.

We further investigated these differences in transfer task difficulty and whether they affected our crucial measure of flexibility transfer, i.e., the LWPS effect, in follow-up exploratory analyses. We conducted a 2 (group: overlap v. nonoverlap)  $\times$  2 (trial type: switch v. repeat)  $\times$  2 (PS: high v. low) repeated-measures ANOVA on the subset of trials involving only the transfer task across both groups. Participants exhibited larger RTs ( $F(1,168) = 52.9, p < .001, \eta^2 = .24$ ) and error rates ( $F(1,168) = 93.9, p < .001, \eta^2 = .36$ ) in the transfer task for the overlap group (face emotion task; RT: 836.9 ms; error rate: 0.21) compared to the nonoverlap group (object origin task; RT: 711.9 ms; error rate: 0.08). Transfer task switch costs were also higher in RTs ( $F(1,168) = 9.1, p = .003, \eta^2 = .05$ ) and error rates ( $F(1,168) = 28.7, p < .001, \eta^2 = .15$ ) for the overlap group (RT: 42.3 ms; error rate: 0.04) compared to the nonoverlap group (RT: 32.4 ms; error rate: 0.10).

These results confirm that the transfer task (face emotion task) was more difficult and suffered from more cross-task interference. However, crucially, there was no interaction between group  $\times$  trial type  $\times$  PS ( $F(1,168) = 0.25, p = .62, \eta^2 = .002$ ), indicating that these differences in task difficulty and cross-task interference did not significantly affect the within-task switch costs changes across different PS conditions.

We also investigated whether the magnitude of the LWPS effect differed in the biased tasks (face gender and age tasks) across the two groups in an exploratory 2 (group: overlap v. nonoverlap)  $\times$  2 (trial type: switch v. repeat)  $\times$  2 (PS: high v. low) repeated measures ANOVA. There were no significant main or interaction effects of group in RTs or error rates. However, the interaction between group  $\times$  trial type  $\times$  PS approached the cut-off for statistical significance ( $F(1,168) = 3.4, p = .067, \eta^2 = .02$ ) in RTs, whereby the LWPS effect was numerically slightly larger in the overlap group ( $\eta^2 = .28$ ) compared to the nonoverlap group ( $\eta^2 = .20$ ). Given the small difference in effect sizes, it is unclear whether this nonsignificant interaction reflects any meaningful differences in the magnitude of the LWPS between groups.

In sum, since there were no significant differences in the LWPS effect across groups for either the transfer task or the biased tasks, our main interpretations of Experiment 3 results remain unaffected by differences in task-specific performance. However, for future investigations, we recommend implementing the same transfer task for both groups for a cleaner design. For example: using two face tasks (age and gender) for the biased tasks and the face emotion task for the transfer task in the overlap group and using two object tasks

(origin and size) for the biased tasks and the same face emotion task for the transfer task in the nonoverlap group.”

## General Discussion

Across three cued task-switching experiments that utilized LWPS manipulations, the current study replicated findings from Siqi-Liu & Egner (2020) and provided additional evidence that flexibility learning generalizes over novel stimuli and physical task-cues. Furthermore, flexibility learning seems to be tied to abstract task-sets rather than physical task cues, and this robust task-specific learning occurred regardless of whether there is stimulus overlap between tasks.

Experiment 1 replicated prior findings of list-wide flexibility adjustments (Bonnin et al., 2011; Dreisbach & Haider, 2006; Dreisbach et al., 2002; Duthoo et al., 2012; Mayr, 2006; Monsell & Mizon, 2006; Siqi-Liu & Egner, 2020) but using trial-unique stimuli, thus providing the most stringent evidence that flexibility adaptations do not rely on learning of any kind of stimulus-level bias. To our knowledge, this is the first time that the LWPS effect has been found with completely trial-unique stimuli. Experiment 2 probed whether flexibility adaptations depended on control associations with abstract task-sets (i.e., the rules that define stimulus-response associations) as suggested by Siqi-Liu & Egner (2020, Exp 3), rather than associations with the arbitrary physical cue stimuli, or the visual features that indicated to participants which task they should perform – in a similar manner that task-irrelevant contextual features have been shown to trigger conflict adaptation effects (Bejjani et al, 2018). We observed flexibility adaptations to PS context in trials using biased task-cues, which occurred more often on switch/repeat trials in the high/low PS blocks, but also in trials using unbiased transfer cues, which appeared equally often on switch and repeat trials, regardless of the PS context.

In conjunction, Experiment 1 and 2 suggest that flexibility adaptations in cued task switching do not depend on control associations with specific physical features of the trial, whether these be features of the task cue or the target stimulus. Rather, flexibility seems to be specifically tied to the categorization rules that are applied to task stimuli. These results support the conclusions from Siqi-Liu and Egner (2020) that flexibility learning in the LWPS protocol is task-specific, and that task-, but not stimulus-level associations are necessary for flexibility adaptations.

Experiment 3 further tested whether task-specific flexibility learning would still occur even when tasks are performed on the same stimuli, with integrated, overlapping features. Across two groups of participants, we compared cross-task transfer of the LWPS effect from two biased face tasks to a third unbiased transfer task, in one group utilizing the same face stimuli, and in the other group using an unbiased object categorization task with a completely non-overlapping stimulus set. In both groups, we found LWPS effects in the two biased tasks, but not in the unbiased transfer task, suggesting that stimulus-feature overlap, or the lack thereof, did not have any effect on task-specific flexibility learning. It is important to note that the lack of a statistically significant LWPS effect in the transfer tasks does not necessarily mean that context-specific switch cost adaptations did not occur, since



nonsignificant findings can also be explained by a lack of power (despite our sample size calculations) or chance. However, Experiment 3 results generally suggest that task-specific flexibility learning is highly robust and is not affected by stimulus-set overlap in between task-sets.

Siqi-Liu & Egner (2020) speculated that task-specific switch readiness adjustments may occur via rapid bottom-up priming of context-appropriate control settings (King et al., 2012), which aid preparatory task-set reconfiguration processes for those frequently switched-to tasks. Evidence that task-sets may specifically define the boundaries of control strategies has also emerged in studies of the congruency sequence effect (CSE), which was eliminated when sensory modalities (for stimuli and distractors) changed, but only when different sensory modalities defined different task sets (Grant et al. (2020)). That is, when the distractor modality predicted the target modality, participants used the distractor information to orient themselves to the modality in which the target stimuli would appear, facilitating the formation of modality-specific task sets. These modality-specific task-sets acted as boundaries for the CSE – when task-sets (rather than simply target or distractor modality) switched, participants abandoned control expectations formed in the previous trial, thus eliminating the CSE (Grant et al. (2020)). Similarly, in task-switching paradigms, task-sets may help participants orient to the relevant stimulus dimension for producing a correct response, providing a strong “context” to which flexibility learning could be bound. In other words, in cued-task switching, task sets may play a key role in determining processing strategies (Egner, 2014; Hazeltine et al., 2011; Schumacher & Hazeltine, 2016) and participants may be encouraged to favor trial-by-trial updating of flexibility based on task-cues over implementing global shifts in their updating threshold across blocks.

However, the current set of experiments only examined flexibility learning in the context of cued task-switching and does not rule out the possibility that cross-task flexibility transfer could occur under other experimental conditions. Specifically, cross-task flexibility transfer was recently demonstrated in a probabilistic version of the Wisconsin Card Sorting Task (Wen et al., in press). There, participants completed the card sorting task in either high (frequent uninstructed rule switches) or low volatility (infrequent rule switches) environments before transitioning to a medium-volatility transfer phase. Using reinforcement learning modeling, authors found that participants exhibited higher learning rates, or faster adaptation to rule changes, in transfer phases that followed exposure to the high-volatility environment. These learning rate adaptations occurred even when novel tasks and stimuli were used in the transfer phase. Wen et al.’s (in press) paradigm may be particularly effective at inducing general changes in flexibility because participants had to voluntarily switch between categorization rules to discover rule switches based on feedback amidst uncertainty. Optimal performance may thus depend on changes in general exploratory behavior because of the lack of explicit task cuing.

In contrast, cued or forced task-switching does not involve any uncertainty regarding the correct rule to implement on each trial, and thus may encourage participants to prioritize maintaining only frequently switched-to task-sets in working memory. In particular, Dreisbach and Fröber (2019) argue that some cases of flexibility modulations may depend on concurrent task activation, rather than the adjustment of a general ‘updating threshold.’

Lowering the updating threshold may increase flexibility in general; on the other hand, adaptations based on concurrent task activation may manifest as increased flexibility between only specific tasks held in working memory. When participants can rely on explicit task cues for the upcoming task, they may rely on task-specific flexibility adaptations via such concurrent task-set activation rather than more effortful adjustments to the general updating threshold which may only be necessary in paradigms without explicit task-cuing such as Wen et al (2022).

Notably, (Ileri-Tayar et al., 2022) also identified cross-task transfer of control learning based on common stimulus features in neutral temporal contexts, providing an interesting example of conditions that may promote global control adaptations. However, the authors' findings may not be directly comparable to the current results since Ileri-Tayar et al. (2022) were interested in persisting conflict adaptation to mostly congruent or mostly incongruent stimuli or stimulus features in neutral transfer lists, while the current study probed flexibility adaptations to list-wide context. Additionally, also investigating contextual adaptations of flexibility, a recent study found that participants adapted to task-irrelevant environmental cues that signal varying demands for switching, although the authors did not test for cross-task transfer (Xu et al., 2022).

Since cross-task transfer appears to vary depending on paradigm design, future research may benefit from investigating other conditions for transfer. For example, flexibility learning transfer has not been tested within a hierarchical task-set organization where multiple simple subordinate tasks contribute to an overarching supraordinate task (Lee et al., 2022). It is possible that, though flexibility-learning does not transfer across tasks conducted in parallel (age and gender face tasks), the same flexibility settings could be shared among subordinate tasks that contribute to the same supraordinate task (e.g., subordinate line orientation or color judgement tasks versus supraordinate line and color judgement tasks).

In conclusion, the current study the findings of task-specific flexibility learning from Siqi-Liu & Egner (2020) under novel circumstances, providing additional evidence that task sets form robust boundaries for flexibility learning in cued task-switching paradigms. Further research is required to determine the conditions of cross-task flexibility transfer in other paradigms, such as ones that involve un-cued task switching or hierarchical task-set organization.

## References

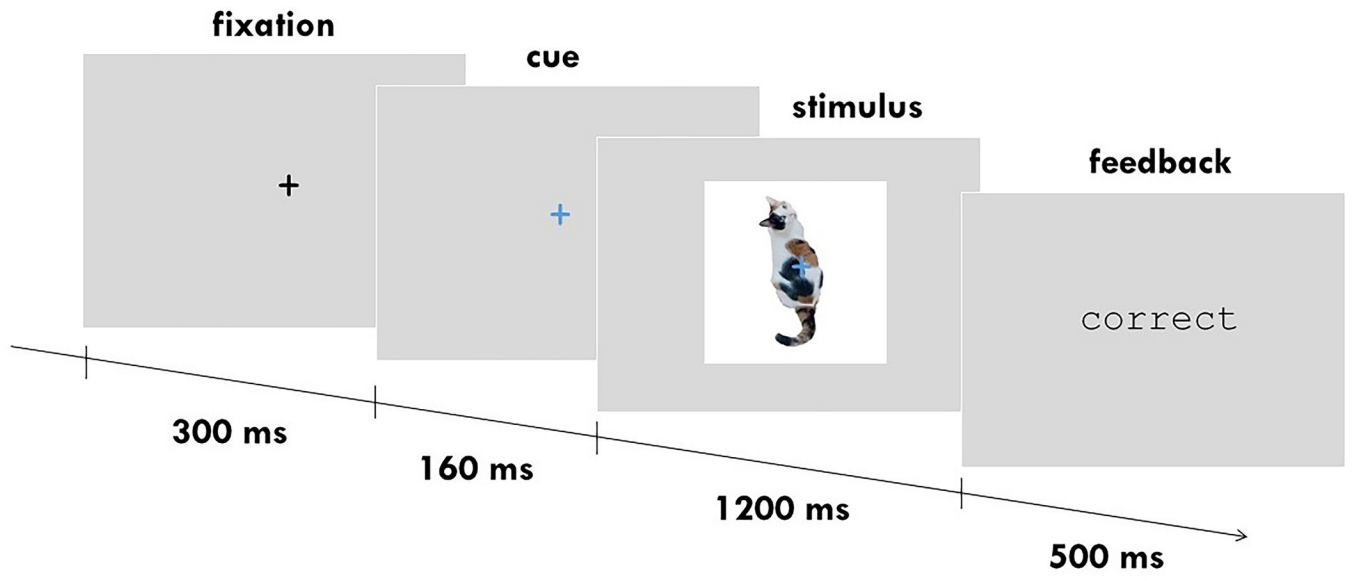
- Bejjani C, Zhang Z, & Egner T (2018). Control by association: Transfer of implicitly primed attentional states across linked stimuli. *Psychon Bull Rev*, 25(2), 617–626. 10.3758/s13423-018-1445-6 [PubMed: 29450789]
- Bonnin CA, Gaonac'h D, & Bouquet CA (2011). Adjustments of task-set control processes: Effect of task switch frequency on task-mixing and task-switching costs. *Journal of Cognitive Psychology*, 23(8), 985–997. 10.1080/20445911.2011.594435
- Bugg JM, & Crump MJC (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. 10.3389/fpsyg.2012.00367

- Bustamante L, Lieder F, Musslick S, Shenhav A, & Cohen J (2021). Learning to Overexert Cognitive Control in a Stroop Task. *Cogn Affect Behav Neurosci*, 21(3), 453–471. 10.3758/s13415-020-00845-x [PubMed: 33409959]
- 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. 10.1177/0963721418800030
- Dreisbach G, & Haider H (2006). Preparatory adjustment of cognitive control in the task switching paradigm. *Psychonomic Bulletin & Review*, 13(2), 334–338. 10.3758/BF03193853 [PubMed: 16893004]
- Dreisbach G, Haider H, & Kluwe RH (2002). Preparatory processes in the task-switching paradigm: Evidence from the use of probability cues. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3), 468–483. 10.1037/0278-7393.28.3.468 [PubMed: 12018499]
- Duthoo W, De Baene W, Wühr P, & Notebaert W (2012). When Predictions Take Control: The Effect of Task Predictions on Task Switching Performance. *Frontiers in Psychology*, 3. 10.3389/fpsyg.2012.00282
- Egner T (2014). Creatures of habit (and control): a multi-level learning perspective on the modulation of congruency effects. *Frontiers in Psychology*, 5. 10.3389/fpsyg.2014.01247
- Faul F, Erdfelder E, Buchner A, & Lang A-G (2009). Statistical power analyses using G\* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, 41(4), 1149–1160. [PubMed: 19897823]
- Grant LD, Cookson SL, & Weissman DH (2020). Task sets serve as boundaries for the congruency sequence effect. *Journal of experimental psychology. Human perception and performance*, 46(8), 798–812. 10.1037/xhp0000750 [PubMed: 32324028]
- Hazeltine E, Lightman E, Schwarb H, & Schumacher EH (2011). The boundaries of sequential modulations: Evidence for set-level control. *Journal of Experimental Psychology: Human Perception and Performance*, 37(6), 1898–1914. 10.1037/a0024662 [PubMed: 21767054]
- Ileri-Tayar M, Moss C, & Bugg JM (2022). Transfer of learned cognitive control settings within and between tasks. *Neurobiology of Learning and Memory*, 196, 107689. [PubMed: 36374800]
- Kazak AE (2018). Editorial: Journal article reporting standards. In (Vol. 73, pp. 1–2). US: American Psychological Association.
- Kiesel A, Steinhauser M, Wendt M, Falkenstein M, Jost K, Philipp AM, & Koch I (2010). Control and interference in task switching—A review. *Psychological Bulletin*, 136(5), 849–874. 10.1037/a0019842 [PubMed: 20804238]
- King JA, Korb FM, & Egner T (2012). Priming of Control: Implicit Contextual Cuing of Top-down Attentional Set. *Journal of Neuroscience*, 32(24), 8192–8200. 10.1523/jneurosci.0934-12.2012 [PubMed: 22699900]
- Lee W-T, Hazeltine E, Jiang J, & Jiang J (2022). Interference and Integration in Hierarchical Task Learning.
- Mayr U (2006). What matters in the cued task-switching paradigm: Tasks or cues? *Psychonomic Bulletin & Review*, 13(5), 794–799. 10.3758/bf03193999 [PubMed: 17328375]
- Monsell S (2003). Task switching. *Trends in Cognitive Sciences*, 7(3), 134–140. 10.1016/S1364-6613(03)00028-7 [PubMed: 12639695]
- Monsell S, & Mizon GA (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. 10.1037/0096-1523.32.3.493 [PubMed: 16822121]
- R Core Team. (2019). R: A language and environment for statistical computing. In R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rogers RD, & Monsell S (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, 124(2), 207–231. 10.1037/0096-3445.124.2.207
- Schumacher EH, & Hazeltine E (2016). Hierarchical task representation: Task files and response selection. *Current Directions in Psychological Science*, 25(6), 449–454.
- Siqi-Liu A, & Egner T (2020). Contextual Adaptation of Cognitive Flexibility is driven by Task- and Item-Level Learning. *Cognitive, Affective, & Behavioral Neuroscience*, 20(4), 757–782. 10.3758/s13415-020-00801-9

- Spinelli G, Perry JR, & Lupker SJ (2019). Adaptation to conflict frequency without contingency and temporal learning: Evidence from the picture-word interference task. *Journal of experimental psychology. Human perception and performance*, 45(8), 995–1014. 10.1037/xhp0000656 [PubMed: 31144859]
- Uddin LQ (2021). Cognitive and behavioural flexibility: neural mechanisms and clinical considerations. *Nature Reviews Neuroscience*, 22(3), 167–179. 10.1038/s41583-021-00428-w [PubMed: 33536614]
- Wang YC, & Egner T (2022). Switching task sets creates event boundaries in memory. *Cognition*, 221, 104992. [PubMed: 34929522]
- Wen T, & Egner T (2022). Retrieval context determines whether event boundaries impair or enhance temporal order memory. *Cognition*, 225, 105145. 10.1016/j.cognition.2022.105145 [PubMed: 35483158]
- Wen T, Geddert RM, Madlon-Kay S, & Egner T (in press). Transfer of learned cognitive flexibility to novel stimuli and task sets. *Psychological Science*. 10.1101/2021.07.21.453253
- Xu S, Simoens J, Verguts T, & Braem S (2022). Learning where to be flexible: Using environmental cues to regulate cognitive control. *PsyArXiv* 10.31234/osf.io/y5h78

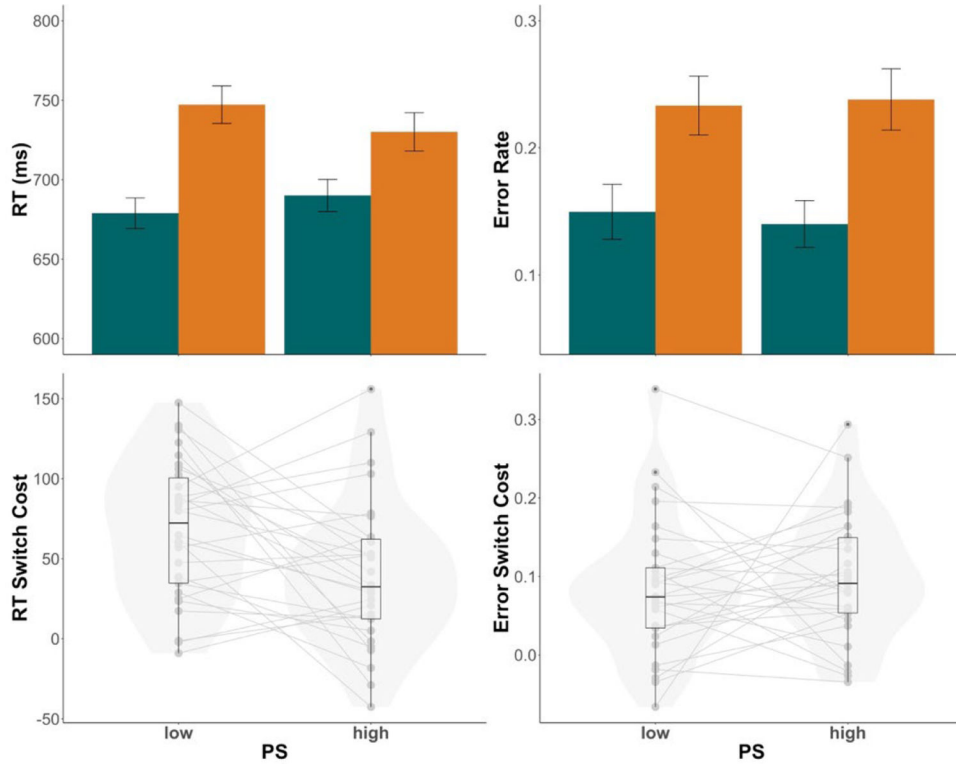
### Public Significance Statement

Different situations in daily life require different levels of cognitive flexibility. For example, a barista at a busy café must rapidly switch between multiple tasks such as taking orders, steaming milk, and pulling shots from the espresso machine, whereas a student studying in a busy café needs to avoid switching attention to the many distractions in his environment to focus on his task. Previous work has suggested that people can match their flexibility (or switch-readiness) to different settings. However, such flexibility learning is “task specific,” such that people only learn to become better at switching to tasks that they have more practice switching to, rather than being more flexible at switching to any task. To further test this hypothesis, in the present study, we investigated conditions where task-specific flexibility learning might not occur. Nonetheless, we repeatedly found that people learned to become more flexible only for the tasks that they practiced switching to more often. These findings are important for understanding the conditions where flexibility learning can or cannot transfer to new situations.



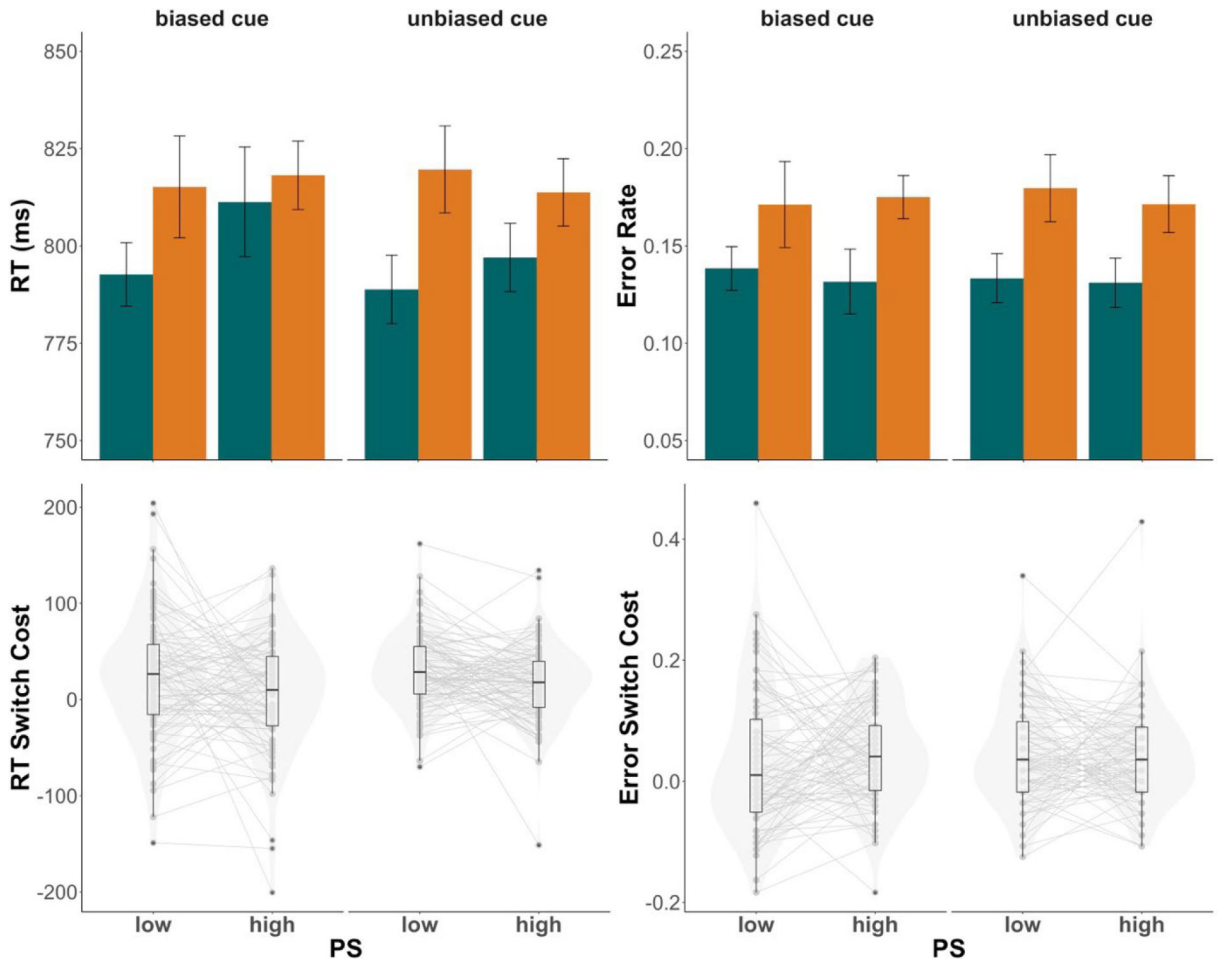
**Figure 1.**  
Example stimulus and trial timing of Experiment 1 protocol.





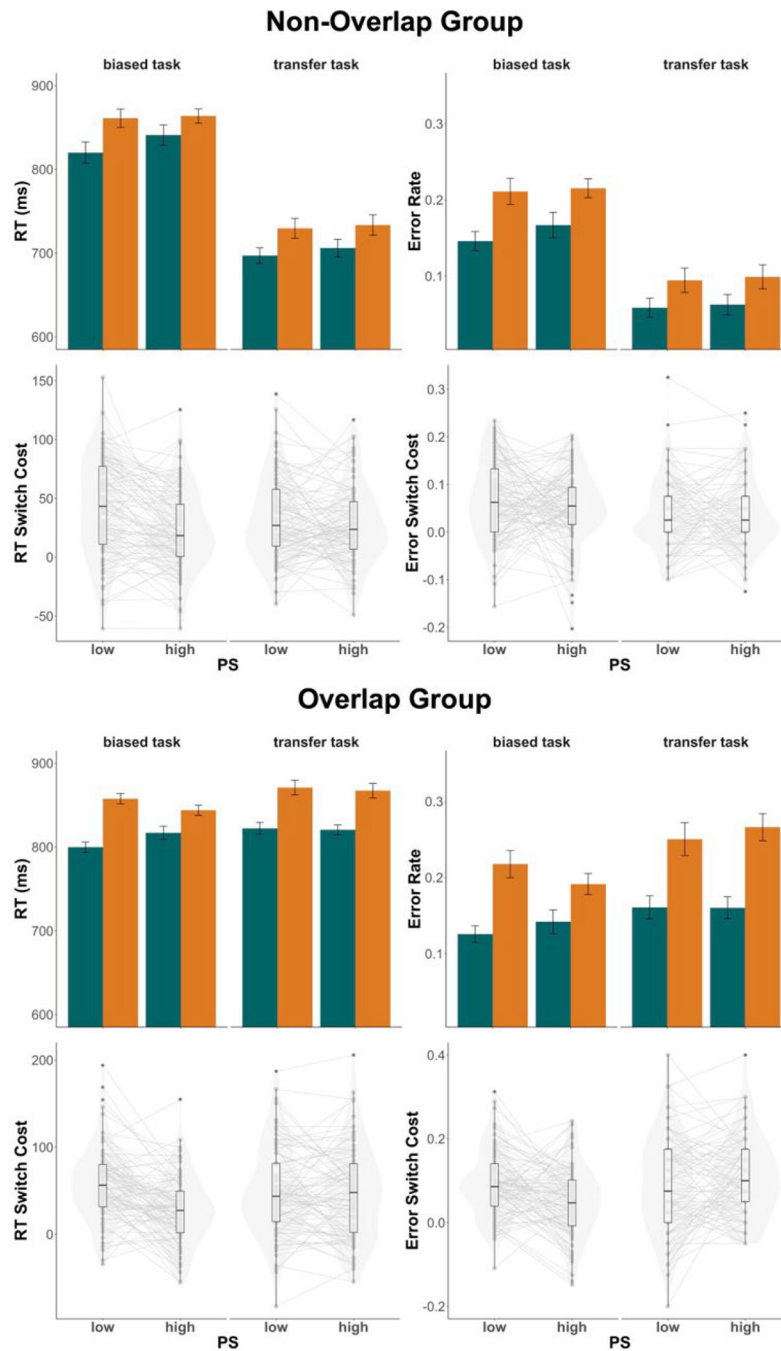
**Figure 2.**

Experiment 1 results. Top panels depict group mean RTs (top left) and error rates (top right) in low (30%) versus high (70%) proportion switch (PS) conditions with switch trials in orange and repeat trials in green. Bottom panels depict switch costs for RTs (bottom left) and error rates (bottom right), which were calculated as switch - repeat. Individual dots represent individual mean switch costs. Violin plots are overlaid to visualize the distribution of individual means. In the box plots, central lines depict group medians, box edges show the interquartile range (IQR), and the length of the whiskers mark  $1.5 \times$  IQR.



**Figure 3.**

Experiment 2 results. Top panels depict group mean RTs (top left) and error rates (top right) in low (30%) versus high (70%) proportion switch (PS) conditions with switch trials means are plotted in orange and repeat trials means in green. Bottom panels depict switch costs for RTs (bottom left) and error rates (bottom right), which were calculated as switch - repeat. Individual dots represent individual mean switch costs. The left side of each panel show trials with biased cues while the right side shows trials with unbiased cues. Violin plots are overlaid to visualize the distribution of individual means. In the box plots, central lines depict group medians, box edges show the interquartile range (IQR), and the length of the whiskers mark  $1.5 \times$  IQR.



**Figure 4.** Experiment 3 results. Top panels of each group condition depict group mean RTs (top left) and error rates (top right) in low (30%) versus high (70%) proportion switch (PS) conditions with switch trials in orange and repeat trials in green. Bottom panels depict switch costs for RTs (bottom left) and error rates (bottom right), which were calculated as switch - repeat. Individual dots represent individual mean switch costs. The left side of each panel show trials with the biased tasks while the right-side shows trials with the transfer task. Violin plots are overlaid to visualize the distribution of individual means. In the box plots, central

lines depict group medians, box edges show the interquartile range (IQR), and the length of the whiskers mark  $1.5 \times \text{IQR}$ .

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript