Experiment 1

**Methods**

*Participants*

The sample consisted of 114 undergraduate students who completed the study for course credit. To exclude participants who did not adequately comply with instructions regarding random task choice, participants who switched tasks on greater than 80% of trials or less than 20% of trials were removed from analyses (*n* = 14). Age and gender characteristics of the final sample are reported in Table 1. The same data from this sample was analyzed in a previous paper focused on the effects of interval length manipulations; the same exclusion criteria and preprocessing stream were used in both papers. All study procedures were approved by the Texas A&M University Institutional Review Board.

*Table 1. Demographics and Task Performance by Task Condition*

|  |  |
| --- | --- |
| Gender % (F/M/O) | 60/40/0 |
| Age | 19.42 (1.52) |
| Accuracy (%) | 94.15 (7.39) |
| Overall reaction time (ms) | 866.40 (177.32) |
| Switch reaction time (ms) | 975.59 (220.21) |
| Repeat reaction time (ms) | 818.46 (151.55) |
| Switch Rate (%) | 46.56 (13.21) |

***Notes:*** *Means and standard deviations are presented for age and each behavioral metric. Gender breakdown is presented as percentage females/percentage males/percentage other (unsure or nonbinary). Behavioral data displayed are calculated after removal of reaction time outliers, post-error trials, and first trials in each block.*

*Paradigm*

Participants performed a modified version of a number Stroop task. Each trial was composed of a task choice phase followed by a task stimulus phase. Task design is displayed in {{Figure 1}}.

In the task choice phase, participants were presented with a ‘?’ in the center of the screen. The ‘?’ indicated that participants were to choose which of two tasks to perform; a physical comparison or a numerical comparison. Participants indicated their choice with a key press (either ‘d’ or ‘f’, task choice mapping counterbalanced across participants) which of the two tasks they chose to perform. In line with {{many voluntary task switching paradigms,}} participants were instructed to choose tasks randomly. Participants were instructed that this meant they should choose each of the two tasks about equally often throughout the experiment. Participants were also instructed that they should repeat the same task as the previous trial and choose to switch to a new task and about equally often. Finally, participants were told not to use a pattern to adhere to these guidelines, but to choose randomly as though they were flipping a coin in their head to decide on each trial. There was no time limit on task choices.

In the task stimulus phase, participants were presented with two numbers that differed in both numerical size and physical size, one number above the fixation cross and one below the fixation cross. If participants had indicated in the choice phase that they chose to perform a numerical comparison, they were to choose the number that was numerically larger in the stimulus phase (ignoring the physical size of the numbers). If the participant had indicated in the choice phase that they chose to perform a physical comparison, they were to choose the number that was physically larger in the stimulus phase (ignoring the numerical value of the numbers). Participants indicated their response in the stimulus phase with a key press (‘j’ for the top number and ‘n’ for the bottom number). Participants had up to 2000 ms to make their choice in the stimulus phase. If the response was incorrect, a message that said ‘Error’ was displayed on the screen. If participants responded correctly, no feedback was presented.

The response-cue interval (RCI) and cue-stimulus interval (CSI) varied between 100 and 1000 ms. Each length was equally likely to occur, and the effects of these interval conditions are not analyzed here. For more information on the effects of these manipulations, please see {{}}. Congruent trials (numerically larger number is also physically larger) and incongruent trials were equally likely. The task consisted of 6 blocks of 65 trials each for a total of 390 trials per participant; task performance lasted about 25 minutes on average.

Prior to the full task, participants completed a practice version of the task, beginning with single task practice blocks, then a shortened version of the full task. If a participant failed to reach 60% accuracy on a given portion of practice, they were required to repeat that portion of practice until the accuracy criterion was reached. To ensure participants understood what was meant by choosing tasks randomly, participants were given feedback after the final practice phase that displayed their task accuracy, switch rate, and percent of trials where they chose each task. If participants switched tasks on less than 20% of trials or greater than 80% of trials, they were asked to repeat that portion of practice. Similarly, if participants chose one of the tasks more than 80% of the time, they had to repeat that portion of practice. Accuracies and RTs are presented along with demographic information in Table 1.

*Analyses*

All analyses were conducted in R version {{# {{cite}}}}. The preprocessing steps and trial inclusion criteria were the same as a previous iteration of the analyses conducted on this dataset involving RCI and CSI lengths{{}}. The first trial of each block (neither a switch trial nor a repeat trial) was removed from analyses. Trials following errors were also removed from analyses to account for post-error slowing. Trials with task RTs less than 200 ms or greater than three standard deviations from the mean task RT were also removed. Because RTs were not normally distributed, they were log transformed for analyses; the transformation yielded an adequately normal distribution. Normality for RTs was determined via visual inspection, as a Shapiro-Wilke test would be overpowered for the number of data points to detect inconsequential deviations from normality {{cite}}.

To test whether overall RT, switch cost RT, or switch rate changed over time, we utilized Bayesian multilevel regressions via the ‘brms’ R package {{cite}} using cumulative trial number as an independent variable. Convergence for all models was confirmed both by visually inspecting chains and by examination of {{R}} statistics (all {{R}}’s ≤ 1.01). An effect was considered significant if the coefficient’s 95% credible interval (CI) did not contain zero.

We first tested for changes in switch cost RT over time by testing an interaction between switch/repeat condition and cumulative trial number on log-transformed RTs. Here, a significant interaction would indicate a change in switch cost RT over time (the difference between switch RT and repeat RT would change as a function of cumulative trial number). In this model, the subject-level main effects of switch/repeat and cumulative trial number were also included in the model. However, a subject-level interaction term was not included as it prevented model convergence and subject-level changes in switch cost RT were not central to the research questions. Next, we tested for changes in overall RT over time using a regression model that included main effects for switch/repeat condition and cumulative trial number, but not an interaction term, as group-level and subject-level IVs. Log-transformed RTs were again the DV in this model.

To examine changes in switch rate over time, we used a logistic regression in which task choice (switch coded as 1 and repeat coded as 0) was the DV and cumulative trial number was a group- and subject-level IV. Here, a significant change over time in the probability of choosing to switch tasks would be indicated by a significant regression coefficient for the cumulative trial number. A positive coefficient would indicate that participants are more likely to switch tasks as the experiment progressed.

Finally, we conducted exploratory follow-up analyses to whether individual differences in switch rate change were related to individual differences in RT change, which might inform a possible mechanism behind the switch rate change. This involved extracting subject-level coefficients for cumulative trial number from the logistic regression in which task choice was the DV and regressing them on subject-level cumulative trial number coefficients from the regression in which log-transformed RT was the DV. Because a visual inspection of the data and a Shapiro-Wilke test indicated that the distribution of individual-level switch rate changes was significantly skewed (*p* < .001), we tested this relationship using a nonparametric Spearman rank-based correlation.

**Results**

The interaction between switch/repeat condition and cumulative trial number did not significantly predict RT at the group level (*b* = -0.000040 ± 0.000045), indicating there was no group-level change in switch cost RT over time. When the interaction term was removed, the main effect of switching was significant (*b* = -0.071 ± 0.009), indicating the presence of a switch cost RT. Further, the main effect of cumulative trial number was significant (*b* = -0.007 ± 0.001), indicating a significant decline in overall RT throughout task performance. Cumulative trial number also significantly predicted task choice in the logistic regression (β = -0.21 ± 0.08). Participants were less likely to choose to switch tasks as cumulative trial number increased, indicating an overall decline in switch rate as the task progressed. Finally, the degree to which an individual’s switch rate declined was associated with the degree to which their RT declined (*ρ* = .23, *p* = .02). Individuals that declined more in RT also tended to exhibit greater declines in switch rate.

**Discussion**

In Experiment 1, we identified the presence of a decline in switch rate as the experiment progressed. While we replicated the decline in RT present in many previous works outside of a task-switching context {{cite}}, we did not find evidence of a decline in switch cost RT.

While the large sample in Experiment 1 was well-suited to detect changes in switch rates or switch costs over time, there was no manipulation meant to examine the mechanism underlying the change in switch rate that was detected. However, exploratory analyses indicated a relationship between RT decline and switch rate decline, which provides a possible avenue for exploring the mechanism. Importantly, the regression in which we measured RT declines also included a coefficient for switching; therefore, the resulting RT decline coefficients control for the effect of switching on RT. This inclusion was crucial, as without it the decline in RT would be confounded with the decline in switch rate.

Because the reduction in RT over time is often interpreted as an effect of practice, fatigue or both {{cite}}, we reasoned that reductions in switch rate over time might be explained by similar mechanisms. Repeating tasks is sometimes considered a marker of the use of the bottom-up Availability Heuristic, as the task performed on the previous trial is likely the most readily available to the participant when they choose a task on the current trial {{cite}}. Therefore, we hypothesized based on the results of Experiment 1 that fatigue might partially explain the decline in switch rate over time.

However, it is not possible from the results of Experiment 1 alone to rule out the influence of task-specific factors in producing the decline in switch rate rather than a more general process such as fatigue or practice effects that would occur in other voluntary task switching paradigms. For example, while somewhat common in task switching literature {{cite}}, the use of a number Stroop in Experiment 1 introduces more conflict than other task switching paradigms; the resulting demands on cognitive load might affect the rate of fatigue, practice, or other changes that occur throughout task performance. Therefore, generalizing the results of Experiment 1 to a different voluntary task switching paradigm would be fruitful in helping to rule out the possibility of task-specific factors as an underlying cause of switch rate declines.

Finally, while we did find a significant reduction in switch rate over time as well as a relationship between switch rate decline and RT decline, the effect sizes were in the small to moderate range (although it should be noted that they were larger than the often-reported decline in RT over time). While our large sample (*n* = 100) meant we were well-equipped to detect the effects, replicating them in a smaller sample would lend greater support to our conclusions in Experiment 1.

**Experiment 2**

In Experiment 2, we analyzed an openly available dataset published in the Experiment 1 section of {{Frober et al.}} (hereafter referred to as Experiment 2 of this manuscript). Participants in Experiment 2 first performed a standard double-registrant voluntary task switching paradigm. However, unlike in Experiment 1, participants then performed the same task with performance-contingent rewards.

Importantly, the dataset provided an opportunity to replicate the results of Experiment 1 in a different, commonly used version of a double registrant voluntary task switching paradigm; Experiment 2 utilizes a letter-digit classification rather than a number Stroop. Therefore, Experiment 2 partially served to replicate the results of Experiment 1 in a smaller sample from a separate population using a slightly altered task. We hypothesized that if switch rate declines generalize to other voluntary task switching designs than the number Stroop paradigm in Experiment 1, a similar decline would be present in the non-rewarded phase of Experiment 2.

Experiment 2 also served to test our hypothesis that fatigue might underlie switch rate declines. We hypothesized that if the decline in switch rate was related to fatigue, the increased motivation caused by the performance-contingent reward might reduce (or eliminate) switch rate declines in the rewarded phase of Experiment 2.

*Methods*