

Reviewer 3

Part One: Question on Abductive Reasoning

Comment 1: The local and global switching costs are determined by comparing the differences between three types of trials: pure block, repeat, and switch trials. Specifically, assuming that the reaction time (RT) for the pure block and switch trials are held constant, then increase the RT for the repeat trial will result in both an increase in global switch cost and a decrease in local switch cost. This suggests that the increase in global switch cost and the decrease in local switch cost are two sides of the same coin, and are equivalent. Notably, in this experiment, the RT of pure block and switch trials can be considered fixed. The pure block trial was only measured once and therefore cannot be altered. While the switch trial was measured twice under alt-run and random conditions, the results were quite similar (1414 vs. 1451 for alt-run vs. random, respectively) and there were likely no significant differences between them. It is worth mentioning that the error rate (ER) for alt-run switch trial (6.12%) was higher than that of the random switch trial (5.17%). Thus, even if the alt-run switch trial exhibited some advantages in RT, it may be due to a trade-off between speed and accuracy.

To further illustrate my argument, consider the following analogy. Tom and Bob are two fully grown adults whose heights are fixed. Tom is short, standing at 160cm, while Bob is taller, measuring 190cm. Mark, on the other hand, is a teenage whose height is constantly changing. A year ago, Mark's height was 175cm, resulting in a 15cm difference in height with both Tom and Bob. However, a year later, Mark has grown to 180cm. As a result, we observe a decrease in the absolute difference between Mark and Bob's heights, while the absolute difference between Mark and Tom's heights has increased. Of course, we don't need to explain the opposite direction of the changes in these two absolute differences because they are both due to Mark's increase in height. If there is anything that requires explanation, it is why Mark has grown taller.

So, returning to our experiment from the issue of Mark's height, applying the same logic, I would argue that it is not necessary to further explain why the global switch cost and the local switch cost are developing in opposite directions. Instead, we only need to explain why the RT of the repeat trial is longer in the alt-run condition compared to the random condition (alt-run repeat vs. random repeat = 1328 vs. 1260). In this article, the author suggests that this may be because in the alt-run condition, participants need to use extra working memory load to remember the task sequence to ensure that they know what task to perform next. However, I have two concerns with this explanation:

1. Obviously, under the alt-run condition, participants need to use additional working memory load to remember the task sequence for both repeat trials and switch trials, but why is it that only the RT for alt-run repeat trials increases compared to the random condition, while the RT for alt-run switch trials and random switch trials show no significant difference? (In fact, the RT for alt-run switch trials is even slightly smaller than that for random switch trials: 1414 vs. 1451; we have discussed this issue earlier in comment 1.) In other words, the increase in working memory load should extend the overall RT for participants under the alt-run condition, not just the RT for alt-run repeat trials. The author needs to add an additional mechanism to further explain why this interaction between trial type (switch, repeat) \times working memory load (alt-run, random) occurs.

Simply blaming everything on working memory load is not a viable solution.

2. I carefully read the author's method section and my understanding is that, regardless of whether the participant is in the alt-run or random condition, the prompt (which is usually called a task cue BTW) always appears above the target stimuli. If that's the case, why do participants need to spend additional working memory load to remember the task sequence? Participants can simply follow the prompt to complete the corresponding task under the alt-run condition. I think this is indeed a difficult issue to address in terms of experimental design. If there were no prompt in the alt-run condition but there were prompt in the random condition, the presence or absence of the prompt would become a systematic confounding variable.

Unfortunately, in this paper, I have further evidence to suggest that participants may not have actually memorized the task sequence under the alt-run condition. According to task reconfiguration theory, knowing in advance that the next trial requires switching to a new task is very helpful in reducing the RT for switch trials. Therefore, if participants in the alt-run condition did devote a portion of their working memory load to remembering the task sequence, they would be able to prepare the up coming task-set 500 ms earlier (since the inter-trial interval in this study is 500 ms) than participants in the random condition and thus reduce their RT for switch trials. However, in reality, there is no significant difference in RT between switch trials under the alt-run and random conditions. Therefore, I boldly speculate that participants did not memorize the task sequence under the alt-run condition.

In summary, in this comment, I offer two suggestions. Firstly, the author does not need to explain why the global switch cost increases and the local switch cost decreases under the alt-run condition because, according to the author's experimental design and results, they are equivalent. Secondly, we suggest that the author focus on explaining why the RT for repeat trials is longer in the alt-run condition compared to the random condition. Moreover, I provide two reasons why attributing this phenomenon solely to the increase in working memory load may not be viable.

Response: Thank you for taking the time to review our manuscript. While we understand your concerns regarding the relationship between local and global switch costs, we disagree with your assumption that RTs are constant between pure and switch trials, as this view oversimplifies the relationship between trial types. The difference between pure and switch trials fluctuates numerically, even if this difference does not reach conventional significance. However, because switch costs reflect difference scores, numerical differences in switch trials can, and in our experiment, do contribute to differential switch costs. Thus, we disagree with your suggestion that we shift the focus of our manuscript entirely from explaining switch costs to focusing on changes in non-switch trials.

As for your concerns regarding our working memory account of switch costs, our initial submission already included an “additional mechanism” beyond working memory load to explain the dissociation between switch costs. As noted on page 20 of our initial submission, increased global costs on alternating runs trials reflect a combination of working memory load *and* participants additionally monitoring their progress through each set of runs. When a switch is detected, participants then must disengage from the current task-set, activate the alternate task-set in working memory, and then re-engage the task, while repeating the monitoring process.

However, participants would be less likely to engage in this type of monitoring in the random block, given the unpredictable nature of task changes. We have updated our description of this account (pg. 11) to convey our position more clearly and clarify the role of monitoring in this process.

Regarding our inclusion of task-cues within both switch blocks, we agree that excluding these prompts from the alternating runs block could introduce potential confounds, as task-cues are necessary to signal changes in the random switch block. However, we note that if participants were relying solely on the prompts in the alternating runs block, we would expect their responses to be slowed relative to what we report. Furthermore, it is likely that as participants progress through the alternating-runs sequence, they become aware of the pattern, resulting in them monitoring their progress to a greater extent relative to the random presentation block where no discernable pattern is present.

Comment 2: Regarding the result that participants' RT in repeat trials is longer in the alt-run condition than in the random condition, I can provide an explanation for the author beside working memory load, which is "task hierarchy". In short, many studies have suggested that based on task sequence, participants might form some high-level task (e.g., Schneider & Logan, 2006; JEP: general; 2015; Psychom Bull Rev). For example, in an aabbaabb sequence, participants might chunk two adjacent tasks together (i.e., aa=A, bb=B). Hence, aabbaabb becomes ABAB. In this case, although we can divide trials into switch and repeat cases from the perspective of low-level tasks (represented by lowercase letters), from the perspective of high-level tasks (represented by uppercase letters), all trials are switch trials. That is to say, under the alt-run condition, all repeat trials are in fact switch trials from a high-level perspective, which increases participants RT in repeat trials due to the high-level task switch. However, for switch trials, participants need to reconfigure the task-set anyways, so it does not cause much impact (note that in the aabbaabb case, the task set that needs to be reconfigured for both high-level and low-level switches is consistent, so high-level task switching won't cause any extra efforts). This perfectly explains why there is not much change in participants' RT in switch trials under the alt-run and random conditions, but there is a significant difference in RT in repeat trials under the alt-run condition compared to the random condition.

Of course, if explained in this way, it completely changes the agenda of the paper. The author may not be able to propose their research hypothesis from the perspective of working memory load.

Response: We appreciate you providing us with an alternate account to explain the changes observed on RTs for non-switch trials between the alternating runs and random switch blocks. However, by adopting this account, we would be substantially altering our original predictions post-hoc. Furthermore, as noted in our response to Comment 1, our working memory load account similarly explains these differences, particularly when accounting for the additional monitoring processes occurring whenever participants complete the alternating runs sequence.

Comment 3: I think the author's agenda in the discussion section should also change dramatically. Specifically, I do not think the results of this study can be well compared with previous studies based on special participants. One reason for this is, of course, as I proposed in

Comment 1 and Comment 2, I do not believe that the results of this experiment are due to the additional working memory (or attentional control) burden on participants caused by remembering task sequences.

Furthermore, even if we step back and assume that the results of this study are indeed due to an increase in working memory load, we should also note that the increase in working memory load caused by remembering task sequences in normal participants and the impairment of working memory function caused by structural changes in the brain in special participants (such as older adults or Alzheimer's patients) are two completely different things. They are incomparable. I suggest that the author should compare the conclusions of this study with those of other studies that also increase the working memory load of normal participants.

Response: In our General Discussion, we compare our findings with studies which assessed both older and younger adults because, as noted on page 7 of our Introduction, bivalent switch tasks like the CVOE have commonly been used to investigate breakdowns in attentional control and working memory processes that arise due to aging. By omitting these studies, we would be ignoring much of the previous literature using the CVOE task. While the underlying mechanisms behind age-related changes in the brain are separate from changes induced via working memory load, both produce similar patterns behaviorally. Thus, we believe that our comparison to studies of older adults is justified within this context.

While a comprehensive review of task switching changes between older and younger adults is beyond the scope of our paper, previous research supports the notion that aging and working memory load similarly affect memory and attentional systems (see Craik, 1982). For example, Castel and Craik (2003) found that young adults completing a memory task under conditions of divided attention performed similarly to healthy older adults who completed the same task at full attention. Furthermore, both groups performed poorly compared to a group of younger adults completing the same task with full attention. While Castel and Craik were primarily interested in recognition of word pairs, we note that their task similarly required participants to engage in working memory processes to successfully complete the task. Thus, taxing these systems would be expected to produce attentional declines, regardless of whether this additional strain is due to aging or working memory load. We have updated our General Discussion (pg. 21) to include these findings.

Comment 4: Based on Comment1 and Comment2, I think there may be some issues with the author's research aim. First of all, in hindsight, I'm not sure why the author chose to study how different task-switching contexts would affect working memory process by changing the task sequence. A simpler way would be to increase the number of tasks. For example, if the participants need to alternate between three tasks, then the global switching cost may be greater than the alternation between two tasks because they need to maintain three task-sets. Secondly, if the author insists on studying the impact of task sequence on working memory process, they can try to compare the global switch costs under complex sequences (e.g. aababb) and simple sequences (e.g. aabbbaa). Because the two conditions have task sequences, the prompt can be completely eliminated in the experiment, which forces the participants to rely on the task sequence in memory to switch tasks. In fact, the author seems to know this, as they write in the discussion section, "Future research may wish to explore this notion by increasing run difficulty,

such as having participants complete longer run sequences (e.g., 4-4) or by varying run lengths in predictable patterns (e.g., 2-3-2-3; 3-2-3-2, etc.)"

It is worth noting that this is not a criticism of the author's experimental design, but rather from the perspective of the research purpose. If I had planned to study how task-switching contexts would affect working memory process, I would not have designed the experiment from the angle of task sequence. If I had wanted to study task sequence, I would not have started from comparing alt-run and random switching. Even if I did not speak in hindsight, from the logical deduction of the entire introduction, I would come to a similar conclusion. Because the author did not convince me that studying how task-switching contexts would affect working memory process is an important issue (in fact, previous researchers have already studied this issue from different perspectives and with so many clever experimental designs). I also cannot understand why the author had to use the method of comparing global and local switching costs to study this issue.

If the editor gives the author a chance to revise the manuscript, one important task for the author is to rewrite the introduction. They may need to use a different story to convince readers that the current experimental design is reasonable.

Response: One goal of this paper was to replicate some of the external uses of task-switching. Task-switching studies often make use of predictive sequencing (i.e., alternating runs of varying lengths). Like a random switch-sequence, a longer predictive sequence should be more demanding on participants, as the additional length would require that participants maintain task sets in working memory longer, especially if using a variable pattern (e.g., A-A-B-B-B-A-A-A-B-B, etc.). However, while these more complicated patterns are useful for investigating the effects of switch difficulty on working memory and attentional controls, they may lack external validity. Indeed, in one's day-to-life, task-switches often occur unpredictably, rather than in a set sequence. Thus, we elected to include the random switch (rather than simply manipulate run length) to improve external validity in a laboratory setting. Given the editor has graciously allowed a revision, we have further clarified this point in our Introduction (pg. 9).

We agree that adding an additional task-set or further complicating the predictive task-sequence could provide useful insight regarding task-switching processes, though we note that this suggestion is beyond the scope of the present study and may lack the external validity afforded by the random switch task. We elaborate on our previous discussion regarding this point on pg. 20.

Part Two: Methodology

Comment 5: The author used general eta square to represent the effect size, instead of the more common partial eta square. I noticed that the author used ANOVA analysis with the ez package in R language, and ez ANOVA gives general eta square by default. However, partial eta square would be a more appropriate measure, and it can also be easily obtained from the results of ezANOVA.

Response: As noted in your comment, we initially reported generalized eta squared as our primary effect size measure for ANOVAs due to our use of the *ez* package. In our revised manuscript, we have converted all generalized eta squared to partial eta squared.

Comment 6: In studies targeting elderly participants or other special populations, we often find that RT has a very strong skewness due to cognitive limitations. For example, elderly participants can have very slow reaction times in some trials. However, we cannot set short response windows for these participants, otherwise their error rates will soar. I speculate that this may be why Huff et al., 2015 used both Vincentile plots and ex-Gaussian analyses, which are methods more suitable for elderly populations. However, for normal young participants, skewness in data is usually not a problem. For example, in my own task-switching studies, if the response window is set to 2 seconds (or even 1.5 seconds if you really want to push your participants), the data will be close to a normal distribution. However, the authors used a self-paced design that is more suitable for elderly participants, and then used the common statistical methods of Vincentile plots and ex-Gaussian analyses for special participant groups. It turns out that the authors may have overthought it. Because the additional analyses did not provide any conclusions that could not be obtained from analyzing only RT and ER data. So much so that the authors did not even mention the distribution of RT in the discussion. I suggest that the authors remove the content on Vincentile plots and ex-Gaussian analyses, which will make the article more concise.

Response: We used a self-paced design because a) this follows the standard procedure in other CVOE studies and b) it was unclear how much of a slowdown the random switching might produce relative to alternating runs. Our inclusion of the distributional analyses was empirically motivated, as it was an open question as to whether these analyses would detect changes that might otherwise be overlooked by traditional mean analyses (i.e., changes in the tail of the distribution vs. changes in the center). The use of a response deadline as you suggest would have prevented any possibility of changes at the tail of the distribution. Moreover, response deadlines may have contributed to other behavioral changes that may have been an artifact of a deadline. For instance, participants may have been rushed during the task which may have increased the number of errors across trial types. Because only correct responses are included in the RT analyses, there would have been fewer trials available, possibly reducing the reliability of point estimates for each trial. While a response deadline may have been helpful in exaggerating errors across trials and switch cost types, we were primarily interested in how latencies would be affected if participants were required to slow responding to maintain accuracy.

Given that our inclusion of both types of distributional analyses are novel when comparing alternating runs and random switch sequences, we have elected to retain them. However, for the sake of concision and simplifying our results section, we have moved these analyses to an Appendix (pg. 31). We hope you find this as a compromise to your suggestion.

Comment 7: I suggest the author to significantly simplify the results section. Firstly, as mentioned in comment 6, the author can consider deleting all content related to vincentile plots and ex-Gaussian analyses. Secondly, for traditional RT and ER analysis, the author only needs to conduct a 2 (alt-run, random) x 3 (pure, repeat, switch) repeated measures ANOVA. All the desired conclusions can be obtained from the post-hoc tests of this ANOVA. As I mentioned in

comments 1 and 2, it is unnecessary to further analyze the interaction between switch cost type (local vs. global) and presentation (alt-run vs. random).

Response: Please see our response to Comment 6 regarding our approach to streamlining the results section.

While we understand the desire to simplify the results section, replacing our current RT and ER analyses with your proposed model would shift the focus of our results to changes in the trial types rather than changes in switch costs which are the primary dependent measures in task switching studies. We note that our focus on switch costs is consistent with the broader task-switching literature, as there is a precedent for focusing on changes in switch costs rather than changes in trial types (see Huff et al., 2015; Kiesel et al., 2010; Koch et al., 2005; Minear & Shah, 2008; Strobach et al., 2012, for examples). Finally, as noted in our response to Comment 1, assuming that the RTs for pure and switch trials are held constant oversimplifies the issue. Thus, we disagree with the notion that analyzing the interaction between switch cost types and presentation sequence is unnecessary, as changes in switch costs may not be readily apparent when only analyzing the trial level data, which is the case in our experiment.